Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

aCB158
.F4
1991

Federal Forecasters Conference/1991

PAPERS & PROCEEDINGS OF THE CONFERENCE

Cosponsored by

Bureau of Labor Statistics • Economic Research Service • Central Intelligence Agency
National Center for Education Statistics • Environmental Protection Agency
Office of Technology Assessment • Bureau of Economic Analysis
Bureau of the Census • Bureau of Health Professions

Washington, DC
September 12, 1991

United States
Department of
Agriculture

National Agricultural Library

Federal Forecasters Conference/1991

U.S. DEPARTMENT OF AGRICULTURE
NATIONAL AGRICULTURAL LIBRARY

MAR 8 1993

CATALOGING PREP.

PAPERS & PROCEEDINGS

Washington, D.C.
September 12, 1991

FFC/91 Organizing Committee

Co-chairmen:

Norman C. Saunders,
Bureau of Labor Statistics

Howard N Fullerton, Jr.
Bureau of Labor Statistics

Committee members:

Zoë Ambargis
Bureau of Economic Analysis

Peter Blair
Office of Technology Assessment

Paul Campbell
Bureau of the Census

Debra Gerald
National Center for Education Statistics

Karen S. Hamrick
Economic Research Service

John Kort
Bureau of Economic Analysis

Walter A. Rosenbaum
Environmental Protection Agency

Herbert Traxler
Bureau of Health Professions

Methodology Center
Central Intelligence Agency

Received By:
Indexing Branch

Table of Contents

Table of Contents (Cont.)

Morning Session—FFC/91

Mr. Saunders: Good morning and welcome to the Fourth Annual Federal Forecasters Conference, FFC/91. I am Norman Saunders. I work with the Bureau of Labor Statistics and I'm co-chairman of the Organizing Committee for this year's Federal Forecasting Conference.

The FFC has grown tremendously this year. Three agencies have been added to the sponsor's list. Registrations have jumped from about 200 last year to 260 this year. And the number of papers presented has doubled this year, presenting you with the rather pleasant quandary of deciding which to attend.

The credit for all of this must go to those agencies that support the FFC and to the representatives of those agencies who have served on the Organizing Committee. I would like to take just a minute now to introduce each of these people to you.

First, I want to thank my co-chairman, also with the Bureau of Labor Statistics, Mr. Howard Fullerton. Representing the Bureau of Economic Analysis, in the Department of Commerce, Zoë Ambargis and John Kort. From the Office of Technology Assessment, U.S.Congress, Peter Blair. Representing the Census Bureau, in the U.S. Department of Commerce, Paul Campbell. Representing the National Center for Education Statistics, U.S. Department of Education, Debra Gerald. From the Economic Research Service, U.S. Department of Agriculture, Karen Hamrick. Representing the Environmental Protection Agency, Walter Rosenbaum. From the Bureau of Health Professions in the U.S. Department of Health and Human Services, Herbert Traxler. Finally, we want to thank the Methodology Center of the Central Intelligence Agency for their financial support.

Now I would like to introduce to you the man who will welcome all of us to this beautiful facility at the U.S. Department of Agriculture.

For the past two years Bruce Gardner has been the Assistant Secretary of Agriculture for Economics. In that role he exercises direction and oversight of USDA's economic and statistical agencies and functions as the Agriculture Department's chief economist. Dr. Gardner has taught at the University of Maryland, Texas A&M University, and North Carolina State University. He has served as Senior Staff Economist at the President's Council of Economic Advisors. Please welcome Dr. Bruce Gardner.

Dr. Gardner: Thanks very much, Norman. I would like to welcome you to the Department of Agriculture for this annual Federal Forecasters Conference.

We all know forecasting data tends to be a thankless task. I'm sure that the professionals at the agencies I deal with— the National Agricultural Statistics Service (NASS), the World Board, the Economic Research Service— know about this thanklessness from my point of view because I've had many occasions to use their forecasts and I hardly ever thank them when they are right. Instead, I ask why they missed foreseeing such and such. And of course, one can almost always ask this question because almost no forecast will be correct; and often they are not even close. With this in mind, I think all of us will find heartening a recent article that appeared in the London ***Financial Times*** of September 5, 1991. The ***Times*** was commenting on a controversy about USDA's cattle-on-feed statistics.

Some in the cattle industry have been questioning the cattle-on-feed inventory estimates because the slaughter of cattle after these estimates were published did not indicate the number of

FFC/91 Organizing Committee: From left to right, Howard Fullerton, BLS; Debra Gerald, NCES; Herbert Traxler, BHPr; Paul Campbell, Census; Zoë Ambargis, BEA; Norman Saunders, BLS; Karen Hamrick, ERS; and John Kort, BEA.

cattle going to market that the number we said were on feed would suggest should have been going to market.

The conclusion was drawn that NASS was overstating the number of cattle on feed. However, in the last few weeks it appears that cattle have begun to come to market in larger numbers and, more importantly, these are heavier cattle. As a result, cattle prices have taken a substantial fall. The steer prices you see quoted at Kansas City are other indicators of prices.

The ***Financial Times*** had an article on September 5 commenting on this situation. They are commenting on the fall in cattle prices, and this is what they say:

> "Why the markets took such a tumble is a matter of controversy. At the center of it is a Denver-based producer's organization called CattleFax. Earlier this year, CattleFax questioned the figures from the U.S. Department of Agriculture on the number of cattle in feedlots, saying that based on returns from its own members it believed there were some 300,000 fewer cattle in lots than the USDA calculated. It suggested that the demand for beef was set to continue and consequently that prices would not lose more than six cents a pound going into the summer when seasonally-plentiful supplies usually bring a price drop.
>
> "The organization turned out to be wrong about cattle numbers and about the prospects for prices, which plummeted. But before this was clear, CattleFax's doubts about the USDA figures had prompted a crisis regarding the government data among industry analysts, particularly in some influential security houses."The result of this was that many in the cattle business were led to believe more optimistic scenarios than reality warranted. Based on these misinterpretations of the facts, cattlemen proceeded to keep their animals in feedlots longer. This was to prove a disastrous mistake."

Now, I found this article heartening in that this is a case where people felt the sting of not believing our numbers when our numbers turned out to be right. This episode prompts me to issue an overdue note of appreciation to all those who provide the projections, outlooks, estimates, guesses that policy-making is absolutely dependent upon.

In addition to that note of appreciation, I have also been impressed, in the statistical agencies, with the extent of effort in trying to improve forecasts. You look at the deviations between what has been forecasted and what actually occurred, have spent a lot of effort trying to find out why these deviations have occurred, and on how to improve things. I am sure that the drive to improve our product is a big part of why you are all here today.

So, with that, I welcome you and hope you have a productive conference. Thank you.

FFC/91 Forecasting Contest Winners: From left to right, Clifford Woodruff, BEA; Betty Su, BLS; Thomas Snyder, NCES; Thomas Hady, ERS (Overall Winner); Sal Corallo, NCES; Patrick Walker, U.S. Courts; and Lawrence Sink, Census.

Mr. Saunders: Thank you, Dr. Gardner. We come now to the moment that at least 47 of you have been waiting for: the announcement of the results of the First Annual Federal Forecasting Contest. To carry out this pleasant task, I would like to introduce our two contest administrators, Debra Gerald and Karen Hamrick.

Ms. Hamrick: For the first time, this year's conference sponsored a forecasting contest, and Debra Gerald and I volunteered to

administer it. What we were looking for when we developed the contest was the Renaissance forecaster, someone who could forecast accurately five very different things. Those things were, first, the August unemployment rate; second, the August bank prime rate; third, the spot oil price at the end of August; fourth, the Air Quality Index at the end of August; and, finally, the Baltimore Orioles win record at the end of August.

I would like to thank everyone who entered. We had a lot of fun with this contest and I hope you did too. Those who received an Honorable Mention are:

Larry Sink, Bureau of the Census
Lyle Spatz, Bureau of Economic Analysis
Betty Su, Bureau of Labor Statistics
Sal Corallo, National Center for Education Statistics
Thomas Snyder, National Center for Education Statistics
Patrick Walker, Administrative Office of the U.S.Courts
Clifford Woodruff, Bureau of Economic Analysis

Ms. Gerald: The winner of the Forecasting Contest is Thomas F. Hady, Economic Research Service. Congratulations!

Following the morning break, a poster with the names of the individuals who received awards, as well as a list of reported values for the forecast items, will be on display. Also, these individuals will be recognized in the proceedings of the Fourth Annual Federal Forecasters Conference scheduled for release later this year.

Congratulations to all of the winners!

Mr. Saunders: The theme that we selected for this year's conference is the "Coordination of Federal Forecasting." Michael Boskin, the Chairman of the President's Council of Economic Advisors, has begun an initiative to more closely coordinate the work of the various statistical agencies and to improve the quality of the statistics they produce.

But forecasters, those who look to the future based on what has happened in the past, make up a relatively small proportion of that statistical community.How does this concept of coordination affect us in our day-to-day work? Indeed, what exactly do we mean by the term "coordination" in this context?

When I was sitting down last night trying to jot down these notes, I was trying to remember when our organizing committee sat around the table and discussed the theme for this year's conference. As I tried to remember exactly what we had said about it, it came to me that everyone around that table had a different idea about what the coordination of federal forecasting meant to them.

And so, rather than trying to define it to you now, I'm going to take the easy way out and introduce to you our keynote speaker, a person who has been in the position many times over the years to formulate her own opinions on this topic.

She has served for four years as the Chief Economist at the U.S. Department of Commerce, at a time when that position included oversight of the work of both the Bureau of the Census and the Bureau of Economic Analysis, as well as coordination of all Executive Branch statistical programs. She has served as Senior Economist to the Joint Economic Committee of the U.S. Congress and she has been with the Council of Economic Advisors.

Her years of work in the public sector have uniquely qualified her to discuss with you this morning our theme, the "Coordination of Federal Forecasting." Ladies and gentlemen, I give you our keynote speaker, Dr. Courtenay Slater.

Dr. Slater: Good morning. I am very pleased to be here. I don't know whether to feel like I'm coming home or whether I'm entering a new world. I am a former federal employee. I did once pretend to some credentials as a forecaster. But the operating environment in those days was quite different than what we have today. I'm thinking here of the late 1960s, about a quarter of a century ago, although from my perspective, it doesn't seem that long.

I was first a very minor participant in the forecasting process of the Council of Economic Advisors and watched what they did, and then I moved to the Joint Economic Committee of Congress, taking along my new forecasting expertise. I thought I would just take a minute to set the scene there.

There were no budget committees, of course. There was no Congressional Budget Office. Those had not been thought of, much less established. It was the Joint Economic Committee that had the principal and just about the only congressional responsibility for reviewing the administration's forecast, for presenting alternatives and in general, tracking the overall state of the economy.

Our operational equipment consisted of some old-fashioned electric calculators. The old fashioned Marchant or Frieden calculators, which I can see most of you never heard of because you're too young. They were at least the size of a very large old-fashioned pre-electric typewriter, at least. They were probably heavier than that. You plugged them in, and you entered some numbers and they would chug away for three or four minutes and do a long division for you. And then, if you wanted another long division, you plugged in some more numbers, and it would chug away very noisily for several more minutes.

That actually is what we had when I first went to the Council of Economic Advisors in 1967, but while I was there they installed the first generation of electronic calculators, a very expensive elaborate system which had a box in one place and lots of cables running all over the building, so that people could have little calculators on their desks. It was a marvelous step forward in terms of convenience, particularly if your main activity was calculating compound growth rates, for which you no longer had to know the formula; you just did them. There is a formula, for those of you who came along later.

So, in any event, at the Joint Economic Committee our forecasts were of necessity judgmental. If we used the back of an envelope and a pencil, that was rigor in the forecast. We did have some very good judgmental forecasters, but it was not the kind of forecasting that gets done today.

In the private sector, econometric models were just emerging from academia and moving into the commercial world. Data Resources was established in 1969, I think it was. And we at the JEC became one of DRI's earliest customers. But our use of their marvelous new econometric model was hampered not only by our lack of expertise but also because the on-line service was frequently interrupted and always when you needed it most, of course. Usually the reason was because they were building the Metro over at Union Station and somebody cut the phone line.

Along with our acquisition of on-line service we also had another triumph over the Senate procurement people. We argued successfully that they really should buy us these little hand-held electronic calculators which had just come on the market. There was great reluctance because they were so small. We could put them in our briefcases; we could take them home. They might get lost, and they cost — you know,they cost a lot of money in those days, $50 or $75, or something. But we finally convinced them that we really needed these little calculators so they let us have

FFC/91 Morning Session Participants: From left to right, Ronald Kutscher, William Butz, Courtenay Slater, and Calvin Kent.

them, although they did come around every week to check the serial numbers and make sure they were still there.

My point in all this reminiscence is that in those days coordination among forecasters was not the problem most on our minds. The community of serious macroeconomic forecasters was small. If we had had a forecaster's contest in those days and awarded ten honorable mentions, everybody would have gotten one.

So, coordination, whether it was within the Government — the coordination among forecasters, that is, because the Government had lots of other coordination problems then and now which are very similar — but among forecasters, whether within the Government or reaching out to academia as well, could be handled informally. Our problem was of obtaining and learning to use the new tools that would let us effectively draw on the still-new science of econometrics to improve our forecast.

Well, of course the world changed very rapidly. By the time I left the Government in 1981, macroeconomic models, with all their attendant benefits and problems, had long been standard operating procedure. They're just taken for granted by forecasters everywhere.

I have drawn my examples from my own background in macroeconomics, but a similarly rapid introduction of model-based forecast projections and simulations took place in other fields in the 1970s, driven by the advances in modeling techniques and the availability of computers, and also by the endless hunger of policy-makers for quantitative information.

The Congressional Budget Office was established and burdened with the requirement to estimate the cost of every piece of legislation that is considered in Congress, and this was one of the forces making for the rapid introduction of microsimulation models for estimating impacts of proposed program changes.

Take another example; an Energy Information Agency was established and charged with making forecasts of energy demand and supply. And there are lots and lots of examples in other fields. So today we can bring together a large and growing group of forecasters and have contests and engage in one very valuable kind of coordination.

Now, I did note that when Norm introduced me he told you I was going to define coordination of forecasting, but I'm going to disappoint him, I guess. There are so many different dimensions or vectors of coordination that are important and ought to be discussed, I can't discuss all of them. We could discuss coordinating in terms of who coordinates with whom, coordinating within your agency, within your department, with other agencies, with Congress, with state and local governments, with other countries, and so forth. All these seem important to me.

We could discuss coordinating in terms of processes—coordinating the inputs, the data, the assumptions, whatever—that go into the forecasts. Coordinating the process, the way it's done. Or coordinating the results, everybody coming up with the same conclusion. Well, you wouldn't want to carry that too far, I suppose. But all of those are questions that offer fertile ground for discussion.

The one I would like to concentrate on this morning is the inputs, particularly the statistical data inputs that go into the forecasting models. Ever since I began to learn anything at all about actual real statistics, I have been concerned by what seemed to me the increasingly disproportionate sophistication of the models relative to the quality of the data fed into them. When I studied econometrics briefly in graduate school, the statistics were just taken for granted. They were just there; you used them and nobody discussed whether they were good, bad or indifferent. But, you don't have to be in the real world very long before you should

discover that statistics, as well as techniques, have margins of error.

Concern about the quality of our federal statistics at the moment is certainly widely shared. The American Economic Association, the National Association of Business Economists, and others have established committees on economic statistics and issued statements of concern. I'm sure most of you are familiar with those so I won't belabor the point.

The administration has responded with its own coordinated interagency review of economic statistics and with a proposed program for improving economic statistics, generally referred to as the Boskin initiative. I'm going to come back to the Boskin initiative in a minute. But first, I want to mention another study that addresses quite specifically the need for better data inputs for modeling.

This is a report that has just come out of the panel set up by the Committee on National Statistics at the National Academy of Sciences. This study was requested and sponsored by the Assistant Secretary for Planning and Evaluation at Health and Human Services and by the Food and Nutrition Service. Its particular subject is microsimulation models — models like the Trim 2 model — that are used to project the costs and impacts of proposed social legislation. The panel's report is called "Improving Information for Social Policy Decisions for Uses of Microsimulation Modeling." Just to introduce a little visual interest, it looks like this. It has a very colorful cover.

The report has just recently been published, and I recommend it to you if you're modelers and are not yet familiar with it. I didn't have anything to do with this report, and I don't get a cut on the sales, but the parts of it that I've had a chance to read do seem to be quite good and quite readable, quite useful.I know simulation isn't quite the same thing as forecasting necessarily, but it is a related art form. And I think the points made in this report about needing high quality statistical inputs are equally valid whether we're talking about forecasting or projecting or simulating.

The report expresses great concern about the quality of the statistical data used in models and about the failure of the Federal Government to maintain adequate levels of investment in statistics or to properly coordinate a statistical program. It makes a number of recommendations for data improvement, many of which are broadly relevant and in a sense, familiar: increase federal investment in the production of relevant high-quality statistical data; strengthen and increase investment in the coordination of federal statistical activities; recognize key interactions among individuals and institutions — there they are talking about relating household data to establishment data, being able to match things up; and increase investment in the evaluation of the quality of survey and administrative data.

These are just a few examples of the more general recommendations. The report goes on to discuss particular databases and what might be done to improve them. But the point that I found striking in the context this morning is that this report, which is addressed broadly to the development and use and validation of microsimulation models, placed such great stress — really led off with — the need to improve the statistical inputs.

As I mentioned, this concern is widely shared. This is one more voice coming along. And the administration's concern has led to the Boskin initiative, which is a comprehensive proposal for multi-year funding of improvements in economic statistics.

Many of you I'm sure are already familiar with the elements in the Boskin initiative, particularly those of you who do macroeconomic forecasting. But for those of you who may not be, they pertain to such things as modernizing the national income accounts, moving to the international U.N. system of national accounts concepts, increasing the coverage of the service sector, separating quality and inflation changes in price data, improving the payroll and household employment survey, improving business establishment list and industrial classification codes. In general, I think it's fair to say the entire package is built around the needs of the national accounts and what is needed to really have a good national accounting system and an accounting system that meets the standards set by the U.N. for all countries.

If you are a macroeconomic forecaster and base your forecasts on the national income accounts, you will appreciate the importance of this initiative and its direct relevance to your work. What about those of you who are forecasting something other than the GNP? Chances are the national income accounts at some point directly or indirectly enter into the base from which you're forecasting or into the assumptions for your forecast.

But even if you do not rely on this particular kind of economic data, you do rely on some kind of data, and it comes from a federal source in most cases. All of us have a community of interest in adequate investment in good statistics.

I want to underscore this point because there have been some people who have lacked enthusiasm for the Boskin initiative because it wasn't addressed to their kind of data. I have run across such people. But what you have an interest in is seeing people at the highest level of government at last concerned about getting better data.

So all of us can and should applaud the initiative taken by the administration in data improvement. And all of us can be concerned right now about the uncertain congressional prospects for these budget requests. These budget requests have, as I understand it, been through the House and Senate, or most of the way through the House and Senate. I don't know if they've all had floor action but they've had committee action and are awaiting conferences.

To sum it up very generally, the parts of the initiative that are in the Commerce Department budget, that is, the Bureau of Economic Analysis and the Bureau of the Census, did not fare very well in the House. Very little was approved of these initiatives. Some part of it was restored in the Senate. In the case of the part of the initiative that is in the Bureau of Labor Statistics, much of it was approved by the House, virtually all of it was cut out by the Senate. So we have opposite but equal forces here.

That's a very rough summary and possibly a little inaccurate around the edges, but that's the general idea. The reasons for cutting things out did not necessarily have anything to do one way or the other with the merit of the request. It was because of the competition for funds for other things. Something had to go. Well, resolution of these differences in conference should occur shortly. I don't think the conferences have been scheduled, but they are coming along soon.

The point that I want to underscore is that the entire forecasting community has an interest in the success of these statistical improvement initiatives, both an interest in the short-run question of whether next year's budget will be approved and an interest in longer run success in terms of sustained statistical improvement efforts that actually result in some statistical improvements. This coordinated initiative to improve economic statistics is unique in the recent history of federal statistical programs.

The 1980s were a time of cutting back on investment in statistics. And even before that, in the 1970s, when we were so busily adopting models, far too little was being done to improve the data. In this case, I know because I was there. It was in the late

1970s that I was at the Commerce Department and was to some extent responsible for trying to get funds to do things to improve the GNP and so forth. It was a very uphill struggle and we didn't get very far.

So the Boskin initiative is the first thing to come along in a number of years in the way of a real serious coordinated effort occupying a prominent place in the President's budget. It is an event. It is not just business as usual. A lot of effort went into putting this proposal together. A lot of high level interagency meetings and a lot of meetings in the private sector among people who are concerned and who had been pushing and urging .

If this effort succeeds, success will breed success. People will be happy and will feel good about this and the door will be open. Opportunities will be there for addressing other statistical needs.

If this effort fails, that is, if the Congress fails to appropriate a useful fraction of the funds requested by the Boskin initiative — clearly it's highly unlikely to appropriate them all; but we can still hope for a useful part of the money — but if that does not occur, data users and the people who need the better statistics are going to face a real challenge keeping this interest alive. You know, you put all of your eggs in the basket of getting this appropriation through and you've got a total rebuff. You don't really feel like going back next year and beating your head against the brick wall.

So that's the short-run challenge, if you want to call it that.I don't know that there's a great deal that most of you in this audience can do to influence the budget process at this point. Probably not, although if you know of anything, do it. But for the rest of you, just cross fingers that some reasonable fraction of the budget request is approved. I think there's a fair chance that it will be. You can never be certain what happens to budgets, but we can hope.

It does seem to me that there is a great deal that federal forecasters can do, should do, might want to do, to help keep the statistical improvement effort going. And undoubtedly you are doing some of these things. But do more, beginning with coordinating among yourselves to identify important data improvement needs. In my experience, one of the reasons it's so hard to get budgets approved for statistical improvements is that the budgets are so hard to understand, and they are so boring, and they are so fragmented. We have, of course, a decentralized statistical system, so that a coordinated effort to improve economic statistics depends on activities, and hence, on funding in a number of agencies — most importantly, the Department of Commerce, and the Bureau of Labor Statistics, but others as well. This, as I don't have to tell, you means going through the budget process in a lot of departments and going through a lot of different committees in Congress. The whole process is fraught with complexities and dangers.

Furthermore, even within a given agency, the budgets are fragmented and they are hard to understand. I invite any of you who have never done so, to try reading the Census Bureau's budget for funds for economic fields. You have 59 different requests from different divisions for different things, even though really they may be all part of a very important coordinated initiative. Often they are, and often it is something that BEA needs desperately to improve the national income accounts. But that big picture doesn't come through that way in the budget, and it certainly doesn't come through that way to the appropriations committees, even if they were concerned about improving the GNP accounts, a need which also may not have been fully brought home to them. Thus it is very, very hard to get the overall picture of what is most important to do to improve economic statistics, or economic and social statistics, or any kind of federal statistics so that federal forecasters can make better forecasts, and policy makers can make better policy decisions. If you start from that big question at the top and try to work down to what it is important to do and why it is really important to do it.

The importance of improved statistics might be better understood. It seems to me it would be very helpful if forecasters got together and said, here's what we really need and why, in terms of what people who are not professional forecasters could understand easily. Now, of course, that's hard to do because everybody needs some different little piece of data. So, before you can give that big picture to people, you've got to build that big picture. You've got to get together and agree on what's most needed and be receptive to each other's needs. So that's one kind of thing that it seems to me forecasters could do. And this group, this conference — I don't know to what extent this is an organization rather than just a meeting— but maybe it's the nucleus of where that kind of thing ought to get started. I leave that to other people to think about.

The next step which I guess I've already touched on is to communicate these needs to people they need to be communicated to: through the channels in the Executive Branch; through the departmental channels; through the Statistical Policy Office at OMB; through discussions with the Council of Economic Advisors; and so forth, so that people at decision-making levels in the administration know that statistical budgets are important and know why they're important. And that, too, is not easy to do because departmental people have other demands on their time and their funds and their interests. They're not interested just in statistics. Even in the Commerce Department where statistics are a fairly big part of the Department, there are some other big parts and there is competition at the departmental level.

The other thing I would suggest is participating in and coordinating with the activities of various professional associations that are concerned— the business economists, the American Economic Association, the American Statistical Association and so forth, and very particularly, the Council of Professional Associations on Federal Statistics, COPAFS, which is a consortium of these professional associations I mentioned and some 15 others, all which meet here in Washington quarterly and discuss problems in federal statistics and attempt to educate the world and educate their professional associations. I'm sure COPAFS would welcome an observer at their meetings from the Conference of Federal Forecasters and would at some point welcome a presentation on what you see as your most serious statistical needs.

Also, I would suggest coordination with the organizations that represent state government employees. People at the state level use federal data in many, many ways in many different organizations within state government who don't always necessarily coordinate with one another. Their needs for data, which are very important needs, also do not get well communicated to decision-makers in Washington. I've been very conscious of one small example of this because I participate in a group that advises the State of Virginia on their revenue forecasting. Actually, I'm not sure how much helpful advice we give, but we do have a very useful experience, for me, of being able to observe the state revenue forecasting process. That process is taken very, very seriously at the state level, at least in Virginia, and I'm sure at other states, because at the state level, if you don't take in the money, you don't spend it. So, in the process of iterating these things, you make the revenue forecast first and then you decide how much you

can spend. If your revenue forecast is off, particularly if it's too optimistic and the money doesn't come in, then the spending has to be cut back. Aid to local school districts, for example, has to be cut back.

This has happened in Virginia because we've been very hard-hit by the recession. This is the stuff that hits you where it hurts at the local level. It is a highly politically sensitive thing. The revenue forecast is taken very seriously. It's done very professionally and they do a very good job. But they are dependent on—for example, payroll employment data from the Bureau of Labor Statistics, data which tend to undergo major revision some months after their first release.

It's very nice to be able to go back and show how your forecast would have been better if you'd had the revised data. But what they need is more accurate data initially. One of the things BLS wants very much to do is improve the payroll employment data, and they have some very good notions of how this can be done. What they need is to get their budget request approved for it. So there are people in the state government who care a lot about these statistics. They're not well-organized either. They don't know who to talk to in Washington.

I'm exaggerating a little bit. There are lots of people from state governments who talk to people in Washington. I'm sure many of you talk to them. But the big message doesn't always get through to the top levels where it counts. So there is another kind of coordination that you might think about.

As I said, I'm sure many of you are doing these things already, and I commend you. My thought is just that at this particular point in time a stepped up effort, and in particular, a more coordinated effort would be very, very productive. We are in a crucial period with respect to keeping interest alive and improving federal statistics. The strong interest in better statistics has emerged after a long dry period, and I do hope this interest will be kept alive. I hope you will do what you can to contribute to that. And that's my homily for the day.

I thank you very much for letting me join you. I think this is the point at which you may ask me questions, if you wish.

Mr. Butz: Bill Butz, Census Bureau. Courtenay, it seems to me that the decade of the '80s cannot be completely accurately characterized by one in which interest and expenditure on federal fiscal has declined. There are other examples, but the two large ones that I would point out are the Decennial Census in which the approved budget requested by the administration and approved by the Congress was more than double in nominal terms for the 1990 Census over whatthe 1980 Census was. There were plenty of opportunities, I think, in the political process for someone to slap the hand and say, no, indeed you can't, or shouldn't have that much.

Another example I would give is the survey of incoming program participation, which was born and funded during this decade. Now, indeed, the initial funding was cut subsequently but is now in the process of being restored. Both of those I think are examples, and I think there are others that suggest that the decade wasn't uniformly one of inattention and the emphasis on federal statistics.

Dr. Slater: To repeat, if you could not hear, Mr. Butz points out that the decade of the '80s was notuniformly one of inattention to statistics. He has gone on to point out that we spent, I believe he said, about twice as much money on the 1990 Census as on the 1980 Census at the end of the decade, and that quite early in the decade the funding for the Survey of Income Program Participation (SIPP), a major new statistical initiative, was approved.

Well, yes, I agree. The decade of the '80s was not uniformly anything. The initial funding for the SIPP was approved in the early 1980s, and that was a very good thing. It was very touch-and-go, whether it would happen. There was quite a struggle about the funding. This was actually one of the historical events that has made me feel that it is really important to get your act together and explain to people in the Congress why the data are important.

Somehow the message did get across in the early 1980's that we were spending billions of dollars on social programs and knew much too little about the characteristics of the people who participate in those programs or who would be eligible to participate in those programs. And the thing we had to do was to get better, more accurate, and more detailed data about households, particularly about household income.

That message did somehow get across. It got across both to a Reagan Administration and to a Democratic Congress. It just barely made it, but it did get across. And that was a major statistical event and an important accomplishment of the '80s.

It was a struggle, and it still is a struggle, and the funding has been somewhat on_again/off again. But still, I think the basic message there is an optimistic one that if you really recognize the need for statistical improvement and you communicate that need, things can happen.

As to why the 1990 Census cost so much more than the 1980 Census, I believe Mr. Butz is on the program later and you can ask him that. I'm glad we spent the money we needed to take this census. It is nonetheless the case, according to any number of studies that have been done, that overall funding for economic statistics — of course, it's hard to measure funding for statistics because they come from 50 or 100 different places and you have to make adjustments for inflation and all that. But studies that have looked at the major agencies that produce economic statistics have all concluded that funding in real terms for the major kinds of economic statistics was declining during the '80s, particularly during the early '80s. It has come back some in recent years, and we can be glad of that.

But the problems didn't start in the '80s. The problems were there in the '70s. My own history doesn't go back before that so I won't try to pinpoint where they started. We tried many times when I was at Commerce Department to get funds in the Commerce budget for efforts to improve the GNP, and we never got much. Sometimes we got as far as OMB, and I think maybe once we got as far as Congress. But we never got as far as an appropriation.

So that is, I guess, one reason why I am so interested in and supportive of the Boskin initiative, because it has gotten this far and I would certainly like to see it go the rest of the way.
I hope that's a sufficiently wordy answer to your question, Bill. Since there are no further questions, thank you very much.

Mr. Saunders: Thank you very much, Dr.Slater. Now we will take a short break and when we come back will have our panel of distinguished Federal statistical managers here to discuss the idea of the coordination of Federal forecasting.

Mr. Fullerton: The appointed time for the second morning session has come. Please be seated. The Organizing Committee asked the three senior officials in the statistical agencies to reflect

on the theme of the Conference on Coordination of Federal Forecasts. To do that, it is best to have senior people because it is at the senior level that coordination of statistical efforts take place.

For today's panel discussion we have three people from three different statistical agencies and, cleverly, we have asked them to speak in alphabetical order. The first is Bill Butz from the Census Bureau. He is the Associate Director for Demographic Programs and he was appointed in January of 1983. His responsibilities include the current population survey, the survey of income and program participation, and other national household surveys. Before he came to the Census Bureau, he was with Rand serving as senior economist and deputy director of labor and population studies. Please proceed.

Mr. Butz: Thank you, Howard. Some of the promotional materials for this Conference advertised that there would be 150 to 200 federal forecasters— it's scary—all in one place. And it is scary for me because I'm not an expert. As a matter of fact, I have no specific credentials to speak to you on the subject of this morning's session except for one thing: for a period of several years, a decade ago, I refused to do any forecasts or projections when people were asking for them. I'm not sure if that qualifies or disqualifies me!

The setting was one in which a colleague and I had been doing research on U.S. period fertility rates and had developed an economic model to explain period movements in age-specific fertility rates in the post-World War II period. And the model seemed to have some explanatory power. During that period of time, in the late '70s and the early '80s, there was a lot of interest in what was going to happen to fertility rates. So, various people tried to draw me out to use the explanatory model, which did pretty well at predicting — or shall I say explaining — turning points in the past, to extrapolate it into the future to predict turning points Would there be a new baby boom or not, and the like? My colleague and I resolutely refused to do so, for which I am very glad, because I know the prediction we would have made would have turned out by now to have been wrong for much of the period.

Now, when I came to the Census Bureau in 1983,there was disappointment in some quarters of the Census Bureau because there were people in the Bureau who had been looking forward for a long time to our population projections being put on a more causal modeling or structural equation approach. Economic variables would thus be added to the explanatory factors in order to inform our projections of mortality, migration, fertility, and the like. These people were certainly disappointed because my experience in my own research had been that the science of economics and econometrics, the results and the theory, were not yet even approaching the state in which a Federal Government agency should enter such variables and such causal models into official predictions.

I thought in the first place that the links between economic variables and demographic variables weren't well enough known. That is, the parameters of such models that relate, for example, women's labor force participation or women's wages or women's education to their fertility behavior weren't very well known yet. Second, if one were going to use such relationships, even if they were well specified and well known and well agreed upon, to forecast the future, one would then of course have to have forecasts of women's wages or women's labor force participation rates or women's education a year hence, ten years hence, fifty years hence. Now, some others in the government disagree and make population projections on the basis of sets of economic variables based on behavioral modeling, whereas we still don't do much of that. So, that's a little background to disqualify myself somewhat, since, in general, I have refused to do the activity that you all do with such daring.

But the topic is coordination and centralization. I don't mean to focus my remarks on projections of population variables, but in fact, some of my comments are probably parochial in this direction and I'm not going to make distinctions. If you hear something that seems to apply, perhaps it does; and if you're in another line of business and it doesn't seem to apply, then it probably doesn't. Well, what in fact would be coordinated or centralized? I think there are five things when we're talking about projections or forecasts, or even predictions.

One is the historical or baseline data on which forecasts are made. Courtenay Slater referred to this morning. It might be possible to coordinate or centralize decisions about such baseline data on the recent past, which of course inform the forecasts. *Second*, the type of model that's used, the class of model might be coordinated centrally. For example, in forecasting demographic variables one can think of three classes of models. One is a demographic model, a parity progression model, for example, or a cohort-component model. A second class of models is time series models, which pay attention, in a mathematical statistical sense— to the relationships among the variables and the residuals. And the third would be a structural equations model, generally thought of as an economic model. Three legs on the stool, if you will. Well, one might centralize or coordinate and say: we're all going to use one of those types of models or we're going to use a particular combination of those types of models. A *third* area that might be coordinated or centralized is the structure and specification of the actual model. One might question the fact that a lot of people in the government agencies are using lots of different kinds of specifications. Some use one set of explanatory variables, predicting variables, and others use another; why don't we centralize that, why don't we standardize that in some sense? *Fourth* would be the assumptions of the models concerning variables and parameters about which there are insufficient data. This insufficiency is always the case. There's something out there that we wish we knew about but we don't, so we make some range of assumptions about them in the model. And the *fifth* category that we might centralize or coordinate has to do with reporting conventions and categories, or timing, of the results of such models.

So those seemed to me to be the five aspects of our work that might be coordinated or centralized. Now, Courtenay introduced another aspect— the conclusions, or the predictions. That hadn't occurred to me because it seemed to me that if one coordinated the historical baseline data, the type of model, the structure of the model and the assumptions, you wouldn't have to coordinate the results because in fact the results would be the same. In fact, I suppose, though, that it would be possible to

centralize or coordinate or standardize the results apart from all of those other four aspects by simply saying: you can go ahead and use whatever methodologies and data you want but when you come to the end, we would like your forecast to look like this. I presume and hope that would not be a serious alternative.

Now, I would maintain that there is no agreement among experts on these five domains of possible centralization and coordination, particularly on the first four. There could be agreement on reporting conventions and categories and timing. But on the first four, I maintain that there is not such agreement among experts, nor can we expect any, because the answers depend on the uses of the data. Let me give you a recent example concerning two federal agencies that I won't name, both of which do population projections for countries around the world. One of them uses parity progression and cohort component methods and produces single year-of-age population numbers, by age and sex and several other categories, and by single years in the projection period. The other agency would like to see conformance among federal agencies and among the contractors who work for them to a relatively sparse, closed, parameterized model with only a few parameters. But this latter model, being more general, would not produce data by single years of age nor by single calendar years; it can't do things that finely.

Therefore, the first agency argues, the sparse model would produce data not so useful in the short-term— that is, the next five or ten years— for people who want to build school buildings and staff school planning, who are interested in military recruitment and the like. Perhaps for the longer term, projections that go beyond 20 years, there wouldn't be much difference in uses between the two. So, I would argue that there is value in diversity. Nevertheless, to be sure, some of our customers, some of the people here in the federal agencies or outside, who use forecasts and projections would like just one set of projections in whatever domain we're talking about; let's say population projections.

Having choices is certainly not costless. The clients, the customers, know this. If there's more than one projection, then they have to look at them, they have to examine them. They have to begin to examine what their uses are for those data. They have to begin to question whether one set of assumptions more nearly matches their expectations about what the world will look like, or at least whether one set of projectors—the analysts, the people—seem to be more reliable, more honest, more communicative about them. So, the clients have to do something, have to examine what they want, if there is more than one set or projections or forecasts. Our responsibility to them then, of course, is to document what we do in those five dimensions that I mentioned before, to document well, and to help them choose which is best for their purposes. But not, I would argue, to make one choice for them, even one set of coordinated choices for them, and especially not to produce for them one averaged or least-common-denominator set of projections designed to satisfy everyone somewhat, and therefore nobody precisely.

On the three dimensions— communication, coordination, and centralization— I would say, yes, we should have better communication than we have between agencies doing projections, between the suppliers of projections and the customers— in some cases, both of those are Federal agencies— concerning these five aspects. I believe that in fact communications are quite good already along most dimensions, not only among us here but among our colleagues and customers in the private sector and state and local agencies as well. But no doubt, they can be better.

Concerning coordination, there is coordination now in population projections, at least in the sense that almost all federal projection programs and forecasting programs that use population projections, coordinate or control these, I believe, to the Census Bureau's national projections. There are one or two exceptions where at least there is still productive communication about the differences in assumptions. I favor more coordination. But I'm wary. I'm wary of standardization and control. And I'm wary of reducing redundancy, which seems to be wasteful but frequently in fact provides a valuable scientific check on assumptions, on deductions, and on implications that the forecasters or projectors draw from their work.

Moving on to centralization: no, I don't think so. I think that experience suggests in a variety of scientific and engineering areas (what we're talking about is a combination of the two if you also include art) that centralization leads quickly to a decrease in innovation, a decrease in experimentation, a decrease in responsiveness to changing customer needs.

Let me end with a brief story. It concerns Kenneth Arrow. Arrow is a Nobel award-winning economist who, as a very young man, was reportedly invited — as the military does — to join a team of young mathematicians in Italy during World War II to do long-range weather forecasting. The team began doing the forecasts and after several months noted that the forecasts had nothing to do with the weather. So, being young and bright and wanting to do something useful, they wrote a letter to the relevant general and said something to this effect: Dear General: you have assigned us to this work. It's been interesting, but we have noticed that our forecasts have nothing to do with the actual weather. And they requested to be reassigned to useful work. Several weeks later a letter came back from the general's adjutant and it said: The general has received your letter. He thanks you. He notes that the enterprise you are involved in is indeed a difficult one. He recognizes, further, that your forecasts have nothing to do with actual subsequent weather patterns. However, you will continue to make weather forecasts because they are needed for planning purposes. So, as you all know, predictions, forecasts, projections, do have value. And the trade of those who do such for a living will always be a lively one. Thank you very much.

Mr. Fullerton: Thank you, Bill Butz. Moving on through the alphabet, our next panelist is Calvin Kent. Dr. Kent is with the Department of Energy. He was appointed by President Bush and confirmed by Senate in the summer of 1990. He's on leave from Baylor University, where he held the Herman Lay Chair in private enterprise and directed the Center for Private Enterprise and the National Center for Entrepreneurship Education. He was elected Mayor of Woodway, Texas, which has concurrent jurisdiction over the regulation of electric utilities and original jurisdiction over gas utilities. Let me turn it over to you, Dr. Kent.

Dr. Kent: It's my pleasure to be with you this morning. I must admit that I'm quite nervous about it for a couple of reasons; actually, a couple of things that happened to me. After it was announced that I would be leaving Texas and coming to Washington, I was asked to give kind of a farewell speech. I was introduced by the person who was in charge of the program, who came up with this definition. He said that an economic forecaster was someone who liked figures but didn't have a good enough personality to be an accountant. I thought, gee, it's good to be leaving Texas. Certainly when I go to Washington, I will find that economic forecasters are, if not venerated, at least revered and respected.

That illusion did not last for long because, while I was going through the process of confirmation and making my visits on the Hill, I got up one morning, listened to WTOP, which assured me that the weather would be bright, beautiful and balmy, and left without an umbrella. As I was leaving from one of my appointments to go to another, Washington was engulfed in a torrential rainstorm and I waited as long as I could. Finally, I had to brave the weather, go out and stand in the pouring rain, and get a taxi. I got into to the taxi wet and smelling like a used sweatsock, and sat down, complained about the quality of the weather forecast and also said that the weather forecasters weren't any better than the people who forecast the outcomes of the football games, at which point the typical Washington cabby turned to me and he says, "yeah, but those people who really don't know what they're doing are those damned economists." So, I've never had any illusions since then about the prestige of what we do here in Washington.

Let me briefly review with you what it is that the Energy Information Administration does and how we are involved in the coordination of forecasting activities, because a great deal of what our work is involves the projections and forecasts of energy production and consumption. We have two aggregate national fuel projections that we release periodically. Our ***Short-Term Energy Outlook***, which we call the STEO, is published on a quarterly basis and it is a 24-month projection of energy trends, looking at all of the major sources of energy within the United States, both production as well as imports, and also taking a look at how that energy will be used. Once a year we produce the Annual Energy Outlook (AEO), which I have discovered is kind of one of those bibles around Washington that just about everybody seems to use and base their own projections upon. It is an annual forecast and it projects sectoral energy trends out to the year 2010. It is our long-term forecast and it does go into more detail than the short-term energy forecast does.

In addition to those two, we also have specific projections that we do on production and distribution of fuels. We have the annual outlooks for electric power, for coal, for oil and gas, and we also publish the ***Bulk Commercial Nuclear Power Prospects for the United States and the World***. All of these provide regional detail and expand upon the AEO forecast and all of these are tied back to the AEO forecast. In addition to these six forecasts, we also publish one international forecast, which is our ***International Energy Outlook***. Here, we boldly project the world markets for energy. We place particular emphasis in this forecast on the world oil market; and we do use the AEO as the basis for our own U.S. forecast. The ***International Energy Outlook*** is also widely used and distributed not only within the United States but, as I found when I travelled elsewhere, it is also widely accepted in other countries as well.

How do we go about putting these forecasts together as well as all of the other forecasts that we do, which are not regular publications of ours. First of all in terms of our economic data, the major external source for both the AEO and the STEO is the DRI model, and we use this for our macro forecasting. We do exercise the model independently and we do modify the DRI model to conform to our own forecasts for energy prices and consumption. In addition to this, we use the data from other agencies. The Federal Reserve supplies us with their thoughts on interest rates; the BEA gives us their data on economic growth and industrial output; BLS provides us with employment and prices and the Census, the population forecast. These DRI forecasts, as we modify them, are then compared to the forecasts from other services such as WEFA and other federal agencies, such as the BLS, the BEA and the CEA. In addition, we collect wide varieties of international data from international agencies, from private sources and from the multinational corporations, as well.

Despite my previous experience, we do rely on the National Oceanographic and Atmospheric Administration to provide us data on heating and cooling degree days by region because, as you can imagine, this is a prime determinant of energy demand and is very, very important to us in our forecast. Our transportation data is provided either under a private contract, which gives us the information on air travel— passenger miles travelled, freight tons miles travelled, scheduled seating on aircraft— or the Department of Transportation, which provides us with highway usage by automobiles and trucks. In addition to this, the Geological Survey provides us with our petroleum resource estimates and the Department of Defense gives us the information on jet fuel supplies. With the exceptions that I have just noted, though, most of our demand models use our own data as the sources of our forecast. We have three principal surveys to supply us with data on the end uses of energy. All of these are three year surveys.

The first of these is our Residential Energy Consumption Survey, or our REC Survey, which we do every three years in which we take a sample of the United States and look in detail at how energy is used within residences in the United States. We do the same in our Manufacturing Energy Consumption Survey, our MEC Survey, for manufacturing industries throughout the United States. And we have a third survey, our Commercial Building Energy Consumption Survey, as well (CBECS). This survey looks at energy consumption and use in the commercial sector. We have proposed and we hope to have funded a Non-Residential Transportation Survey (NRTES) because now we do not have any data on how energy is used in the transportation sector, other than by households, which is 20 percent of total U.S. energy use. Additional information used by us on production, prices and consumption is obtained from the fuel surveys that EIA does periodically on a weekly and a monthly basis.

We do find that our forecasts are extensively used by others. That is why we consider it absolutely important that what we do be very transparent so that the users understand how we make the sausage that we're feeding to them. We make our

forecasts available not only through our printed publications — and we'll have some of these examples over here after lunch — but also we now have an EPUB system so you can obtain our data in that manner electronically, if you wish. Our data and our forecasts were the basis for the President's National Energy Strategy. They are also the basis for the International Energy Agency's quarterly forecast. The Council of Economic Advisers uses our forecasts in the Economic Report to the President. The Bureau of Labor Statistics uses them in the Occupational Outlook. The Environmental Protection Agency uses them in their studies of energy and the environment. And even the Department of Health and Human Services uses them for their Low-Income Housing Energy Assistance Program. In addition to this, we find that other agencies are continually using our forecasts. In fact, last year, a poll that I took in the preparation of these remarks indicated that EIA had done 160 forecasts for other federal agencies or for Congress in addition to the work that we did within the Department of Energy responding to their request for forecasts.

Another major area of forecasting for EIA is contingency forecasting. We wish to be prepared to supply timely and accurate data and forecasts if we face contingencies such as the Gulf crisis or, as what could have been an even more severe situation, the recent events in Russia. Simply, what we do in contingency forecasting is produce, in cooperation with the other agencies that are involved in the Federal contingency process, a Blue Book. That Blue Book is classified and it is updated twice a year. It is maintained through a Database and Projections Working Group which consists of representatives from a wide variety of agencies including the CIA, the FBI, the DOT, OMB, Department of Defense, Commerce, Treasury, the National Security Council, the Council of Economic Advisers, the Commodity Futures Trading Administration, FEMA, and others. We do provide them timely information that can be used not only in real world situations but in the simulations that are done by emergency and contingency groups, including the Department of Defense, the Federal Emergency Management Administration, and others as they plan their various activities.

Since I have given you a view of our forecasting efforts and how we relate to other federal agencies, let me just talk about the issue of coordination, which is why we are here today. I am reminded of the quotation that is attributed to the English playwright and author, Oscar Wilde, who said that if you took all economic forecasters and laid them end to end, they would never reach a conclusion. And I think that may indeed be the case, but I don't think that's bad. As I was listening to Bill's comments, I was reminded of the religious tradition in which, if you agree with the preacher, you jump and shout "Amen." I felt like doing that through most of his speech simply because I share his view that the current decentralized approach is the appropriate approach for us to follow. Indeed, as I surveyed my own people, they convinced me of the fact that the system, even though it is decentralized, is working well and we are getting the data that we need and we feel that other agencies are receiving from us the type of data that they find valuable and useful. That does not mean though that there are not other things that might be accomplished. Certainly, there has been renewed interest in energy forecasting in other agencies and this is creating the possibility of disparities and duplication, if they would begin to develop their own energy models.

What this simply means is that EIA must develop models that are both useful and usable by others. And indeed, we are in the process of reorganizing the agency so that we can accomplish that goal. And as we accomplish this goal of trying to have models that others can use, we are going to have to have more contact and more communication among potential users to make sure that our product contains the information that they need and that they can indeed run our models and employ our forecast in their own analysis. We are developing a National Energy Modeling System, which will enhanced have mid- and long-term forecasting capability. The principal reason I was brought to Washington was to develop the NEMS. I was given a very simple task by the White House, and that was to develop the world's best energy forecasting model. I'm not sure if we will achieve that, but certainly we will not confess it if we don't. The format, of course, for the National Energy Modeling System is that it is going to be modular, it is going to be transparent, and it is going to be flexible. Our goal is to maximize its usefulness to others because we do not see ourselves as the sole users of this modeling system that we are developing. We are going to need to provide meaningful forecasts and we are going to need to provide even newer and better ways of transmitting these to our user groups.

I think that it is important to note that we feel that most coordination can be accomplished through informal methods at the appropriate staff level. We think that we need to continue to foster and expand these informal links, but there will probably be areas where we will need to develop more formal exchanges and we look forward to working with all of you as we do that. That concludes my comments.

Mr. Fullerton: Thank you. Our third panelist is Ron Kutscher, Associate Commissioner for Employment Projections. He is currently the Director of the program at the Bureau of Labor Statistics that develops medium-term projections of U.S. economy, covering gross national product, industry output and productivity, industry and occupational employment demand, and —not least, the demographic composition of the workforce— and that prepares the ***Occupational Outlook Handbook*** and other career guidance and training material.

He has represented the U.S. Government at international conferences on labor markets in Hungary, West Berlin, South Korea, and served on the Secondary School Education for the Changing Workplace Panel established by the National Academy of Science.

Mr. Kutscher: Thank you, Howard. Courtenay Slater's comments started me reminiscing a little bit and made me realize that some time next spring will be my 30th anniversary as a federal forecaster. Perhaps one of my purposes on the panel is so all of you can see what happens to you if you stay in this business for 30 years. Some of us weather that storm, while some of us don't weather it too well. I'm here as an example. I'll leave you to decide whether it's a good or bad example.

The topic selected by the conference is coordination. Why do we need such a conference? Well, a number of speakers, starting with Courtenay, have talked about why we need coordination. The reason is, because we have a very decentralized statistical system in the United States. Each part of that system is charged with certain subjects, methodology, data, or projection/forecasts. So, it puts a lot of burden on those that are working in the various component parts to coordinate with other groups. It's interesting to take just one moment to look at this from an international perspective. In March or April, Bill Butz and I were in Budapest representing the U.S. statistical system at a Conference of European Statisticians. It's very interesting to try to explain to someone in some other part of the world the U.S. statistical system. After about a half an hour, with fog glazing their eyes, you realize that they don't understand how we do things, let alone why we do it that way. That confusion occurs because in most other parts of the world, with the possible exception of Switzerland, the statistical work of the country is done in one centralized organization. So, a forecasting conference in any of these countries would not be meaningful to deal with the subject of coordination because it would be an intra-agency subject and, consequently, you would never have an open conference discussing it.

It's also interesting to me to look briefly as to how those other international statistical agencies carry out their forecasting duties because they do not uniformly carry those duties out the same way. Most statistical organizations take on some projection or forecasting duties, which usually begins with projections of population. Other agencies will do almost all of the policy forecasting for the government. For example, in Norway almost all of the economic policy forecasting for that government is done by their central statistical office, including a new environmental model that they're working on, which I have briefly looked at and found very interesting. So I think you can see the broad range of how forecasting is carried out in other industrial countries. The type of work we do in the Bureau of Labor Statistics, for example, in Canada is carried out in the Department of Labor, not in their Central Statistical Office. So, where and how projections or forecasts are carried out can vary considerably across countries of the world.

Back to decentralization and the need for coordination, I think the previous speakers have dealt well with the fact that we have coordination and I want to make some comments on that. I also want to make additional comments that, at times this need for coordination tries your soul because you do not have control over the data inputs or other forecasts you need for your work. Consequently, it can be very difficult for managers to know exactly when you're going to get the data you need or the forecast you need. I'll have some further observations on that later.

Let me divide my comments into coordination of forecasts prepared by others which we use, coordination of the data we need for our projections, and then make some observation on policy coordination. I believe, in terms of coordination of other forecasts, that I need for our work, I find considerable high-quality coordination. We need population forecasts for our demographic and economic projections. Those projections are prepared by the Bureau of the Census. Now, about like a year and a half ago, we relayed to the Bureau of the Census that we needed a new set of population projections for the current round of BLS projections. Well, on one hand I want to chastise Bill Butz because he didn't provide me those forecasts I needed so he let me down. On the other hand, I want to compliment him because he did so in a very straightforward manner and gave me substantial warning so that I had time to adjust to the fact that I was not going to get the population projections that I needed. One of the themes that I'm going to dwell on is that timing is very important in coordination. Had I not known up-to-the-last minute that I was not going to get these population projections, it would have been very difficult for me to adjust to that.

In our work with the Energy Information Administration, Dr. Kent has related how we use the data they prepare on energy supply, demand and prices, which we need in our model and which we do not have the expertise that they would have Thus, it's very helpful to us. At the same time, in the course of developing our projections, we produce material that are useful to them for their projections. Over the years we've worked out a very good working relationship with the Energy Information Administration in the sharing of projections so that both of our publication and forecast needs can be met. I believe this is an illustration of a very good coordination. There also are many other agencies that provide forecasts that are useful to us. The National Center for Education Statistics (NCES) do teacher supply/demand projections, they also do enrollment projections and projections of degrees, both of which are useful to us, particularly in the preparation of our occupational projections. The provision of these by NCES relieves us of needing to treat those subjects and we can concentrate on areas where I think we have a comparative advantage. Very recently, we've also worked with the Bureau of Health Professions, and we're now working with the National Science Foundation in terms of coordinating some of the projections of scientists and engineers that BLS are preparing with some that they have prepared so again it is clear that a fair amount of coordination of forecasts is going on all the time. But timing is the key. We need to make sure that each of our projections or forecasts are done in a way that meet the needs of the other agencies to the maximum extent possible.

The second aspect that I want to deal with is the very intensive use of data other than forecasts. A comprehensive projection system like BLS' or like the Energy Information Administration's forecasting model is a very large user of data. Some of that data is collected inside your own agency. In our own case very important parts of it are not generated inside our agency. Here is where the coordination problem can rise its sometimes ugly head. You don't have control over when each of these data sets will be produced. If you use multiple data sets, you need to plan carefully when each of these will come into being. Many times the problems of timing and of coordination can be quite large. In cases like the "economic censuses," the Bureau of the Census has a regularly planned program that we can anticipate the data availability that we need for our projection model uses. This is very helpful to us in carrying out our own projections work. In other cases, perhaps the latest example that's caused us the most

problem is the preparation of the input/output tables by the Bureau of Economic Analysis. The delay in publishing the 1982 Input/Output Table caused us to do a lot of additional work in order to have an input/output model in our long-term projections that we prepare regularly. Consequently just as in the coordination of forecasts, the providing of basic input data can be extremely important, and, the timing of that through the setting of publication dates or release and adhering to those is very important to users, whether that user is another federal agency, a federal forecaster, or an outside user. I think all of us need to develop regular schedules, and adhere to them as closely as we possibly can.

One of the other elements in our projections mirrors closely what EIA does that Dr. Kent has just described. That is, we need to provide for all users as transparent a look at what we do as is possible. We do this by describing the assumptions, describe and publish the model, and provide the results in many different formats for many different users. The final element that we have developed in our projections program at the Bureau of Labor Statistics is we evaluate the results once the date of the projections has come. Further, we publish the results of that evaluation in the same publication where we publish our regular projections. So users can look at our accuracy and compare it with weather forecasts or other degrees of accuracy to see whether the Bureau's projections are equal to, better than, or worse than others.

Let me just make a comment on policy coordination because that's a subject that can also cause considerable difficulty. What is the role of policy coordination in developing forecasts or projections. How do you coordinate that? Let me use a very old historical example to talk about the problem of policy coordination so that there isn't anyone that will be embarrassed by this example. The example goes back to the early days of the economic projection system that was developed in the Bureau of Labor Statistics. That was started, as I mentioned, in the early 1960s. When we began to put the project together, there arose a conflict between the Council of Economic Advisors and the Department of Labor because the policy goals of those two weren't always exactly in tune. Namely, the policy of the Department of Labor was to lower the unemployment rate as low as possible. The Council of Economic Advisors, on the other hand, was interested in pursuing economic policies that might achieve long-term stable economic rate of growth with stable prices and a low unemployment rate. So they had goals that did not match in all ways those of the Department of Labor. Well, the project that I started my federal forecasting career on, which at that time was known as the Economic Growth Project was nearly terminated over this dispute over policy goals, not over technical considerations. Somewhere in my files, is an accord signed between the Secretary of Labor and the Chairman of the Council of Economic Advisors. This accord agreed that the BLS projections would develop several alternative projections, one of which would have very low unemployment rates (consistent with the Department of Labor goals) and one of which would have higher rates of economic growth (consistent with the Council of Economic Advisors goals). Consequently, the BLS projections could look at the ramifications of both a low unemployment rate and a higher rate of economic growth. Our projections to date nearly always explore such ranges in policy alternatives. Clearly the issue of policy coordination can also be a very important aspect of whether one can successfully complete a set of economic or other types of projections.

Let me close with an admonition, and that is, know your users. We are in a serious discussion in the Bureau of Labor Statistics now over total quality management. One of the principals of that is, you need to know your customers. But more importantly, I believe is you need to know how your customers use your data. Let me close with two illustrations of how customers use our data. Several years ago I got a call from someone who said, "We're very interested in your forecast of the growth of the demand for engineers because we are considering whether or not to expand our engineering school." And then there was a pause on the other end of the telephone line and that speaker said, "Oh, hell, I'm going to tell you the truth. They've already decided to expand the engineering school; they just want me to collect some data to justify the decision they've already arrived at." (Laughter.) The second illustration I'll use is an unanswered letter that's in my office right now. Someone is saying, "Please tell me what occupation is best to go into if you have only three minor convictions for possession of marijuana, terrorist activities and simple assault." The address is a cellblock in some unidentified prison. So you need to make sure that you know your users and you know how your users are going to use your data. Thank you very much.

Mr. Fullerton: Thank you, Ron. When I hear panel discussions I want to know if the panel members have comments or responses to the other members' presentations. Do any of you wish to respond?

Mr. Butz: Well, I can't think of anything that was said that I disagree with. But I want to respond only to Oscar Wilde's comment about the economists laid end to end. Somebody else said that if all the economic forecasters in the world were laid end to end, it would be a good thing.

Mr. Fullerton: Well, another variant to that is if everybody who went to sleep in church were laid end to end, they would be much more comfortable. Do we have questions from the floor of our panelists? Mr. Andreassen?

Mr. Andreassen: We've heard a lot about the degradation of the quality of statistics. Do you feel there is any veracity to that? Have you felt there has been an improvement over the last ten years maybe, or is it basicallythe same?

Mr. Kutscher: Well, I think it's not easy to give a simple answer to that question. In part, if you listened to Courtenay's response to Bill Butz's earlier question, you can give illustrations of data sets that have had very pronounced improvements over the last one to two decades. It's also equally easy to find data sets that have deteriorated, either through less funding or the need to incorporate methodological improvements. I suspect that much of what we hear about the deterioration of federal statistics is in part based on the fact that the demand for government data may be increasing at

a faster rate than the supply is able to meet and, in particular, more complex decisions are made based on government data; therefore, pressure is put on the data set that maybe wasn't even intended when these data sets was collected.

Dr. Kent: I would comment by saying that the biggest single problem that we face is the increasing demands that are being placed upon us for more and more detailed data. That's what is beginning to bury us, particularly when it's not adequately funded to go along with the request. It seems that no matter how much data we produce, it's never quite enough and we never quite have that data set that somebody is just desperate for in what we produce. There's been a slight tendency — not a slight tendency, there's been a strong tendency — to legislate that certain reports will be produced and certain data will be collected without the funding that is necessary to do that. I think that is probably the greatest single threat right now to data quality, at least so far as we are concerned — the additional demands that are being placed upon us to produce more and more detailed information without subsequently being willing to pay the price tag that is associated with that request.

Mr. Butz: In the small domain that I know something about, I don't think there is any question but that the quality and value of the population-related projections that the Census Bureau people do have improved markedly over this decade. My impression is that part of that improvement comes from an aggressive attention to the customers through the Federal Projections Cooperative and through other mechanisms. I think that part of that improvement also comes from an increased participation of our professional staff in professional publications, professional meetings, peer review, and international colloquia where people can exchange views. And my feeling is that a major part of it comes from an improvement in methods. If any of it comes from an improvement in the raw data, I'm not aware of that. I'm looking at John or Paul or Larry. That may be the case, but I think much of the actual improvement comes from a refinement of methods and an incorporation of new methods to deal with the data as they are.

Mr. Fullerton: Dr. Johnston.

Dr. Johnston: Everybody starts with the population projections and many of the variables that other agencies project might themselves affect either migration or fertility Has anyone thought of starting with something other than the population projections?

Mr. Kutscher: I guess in part the answer to your question, Dennis, is that I'm not sure we have, but I think we've made improvements in the methodology which allows some significant increase in the type of feedback that I think you desire. I'll just illustrate that within the type of projections that I do which has labor force, economic projections that's done with a macro-econometric model not unlike what EIA does. Then we have a detailed input/output model that yields industry output and employment projections. Next, we go through an industry occupational matrix to prepare occupational projections. Well, I just described that in a way that's very sequential and in which each step depends on the previous step. In the way in which we now do these projections at BLS, there is a lot more feedback in that loop and nothing in that system is set until we are completely satisfied with everything in the system. In other words, when reach the stage in the projection of occupations, someone will have done considerable work analyzing health occupations; someone working on the economic side will have done considerable research about how much spending there will be on health care by consumers and by government. These two can be inconsistent, so that we analyze these inconsistencies and at least bring these into compatibility inside our own projections. Now, the problem is that in the decentralized statistical system, it's not possible for me or any other agency to sit down and do simultaneous coordination with EIA, Census, BEA, and with all the other groups that might be desirable. But I think techniques and technology has allowed us to do more of that type of feedback than what we used to do where it was totally a sequential set of projections.

Mr. Butz: Well, Dennis, there's no question that fertility patterns, migration patterns, and mortality patterns don't arise out of nowhere. They come from behavioral as well as random responses on the part of individuals as they are influenced by their surroundings. Indeed, there are a great many people — I was one of them before I came to the government, as I said at the beginning of my remarks — who do research on that and do try to relate, in my case, economic explanatory variables to demographic dependent variables. Explaining the past is relatively easy. Predicting the future requires, as I said, not only that you understand something about the linkages, have parameter estimates that, if you don't believe, at least you think are useful. But also, that you are able to project or predict the future of those explanatory variables. I maintain that it is more difficult five years in the future to know what the GNP will be, what women's wages will be, what differential unemployment rates in different cities which are relevant for migration will be, than it is alternatively to project those demographic variables on the basis of the recent past using relatively straightforward methods. Others disagree. The Bureau of Economic Analysis does some population projections in which, as I understand it, they do use some explanatory economic variables across different regions of the country to induce predicted migration rates. So there are some alternative ways to do that. But I think primarily this is done still in a research mode by private researchers.

Dr. Kent: Just let me make one comment that so far as we are concerned in EIA, one of our basic objectives in all of our forecasts is to achieve an energy balance in which supply and demand are equilibrated to each other. This can only be accomplished through a highly iterative process in which we continually modify our projections based on what our previous data told us and then how that data was modified. Eventually you have to reach a balance. Certainly, we have to take into consideration the population figures. If you're concerned about households, you've got to be concerned about household formations, you've got to be concerned about the standard of living the people are going to have

five, ten, thirty years into the future. But the process is one in which you are continually going back and reinvestigating the data and modifying what your previous assumptions were. And that system I think has served us very well. So I'm not really sure anymore where our process starts. I'm sure where it ends, but I don't know where it starts.

Mr. Fullerton: We have two questions over there. Yes?

Question: The assumption has been made explicitly about the relationship between forecasts on the one hand and policy decisions such as on the other hand. In particular, the reasoning seems to go like this: our forecasts project the need in five or ten years a specified number of special resources or abilities that imply needs that require funds to be appropriated. Therefore, Congress should appropriate these funds. It seems to me this type of reasoning does involve to some extent (inaudible) and I wondered shouldn't we really question this approach because the projections themselves are contingent upon the decision to be made, how many facilities would be available. Ratherthan saying how do you follow from the projection, the projection showed different theories such as (inaudible) contingent upon assumptions about available (inaudible). It would seem to me that the business of relating forecasts, a picture for policy options, (inaudible) requires careful scrutiny of what we think the role of human actions (inaudible).

Mr. Fullerton: You wanted the panel to reflect on the need for different forecasts that would reflect different policy alternatives, rather than just simply have one forecast that maybe specified what the agency hoped would happen? Is that a fair summary?

Question: What I'm trying to say is that it's problematic to say we should bring actions now upon a projection of future conditions (inaudible).

Mr. Fullerton: It's problematic that one should take action based on a projection, rather than a series of projections that take into account various policy alternatives.

Mr. Kutscher: I think in my remarks I threw out a couple of thoughts and I didn't fully elaborate on those thoughts. Thought number one is, there is probably a difference between a technical set of projections and a set of policy simulations. If you asked me what we do, I would say technically we try to provide the best set of projections we can given the data and resources we have available. A policy forecast or simulation can be thought of as turned around the other way and asked, given this policy, what will the end result be? If I understand you correctly, I think you're arguing for what I heard this panel suggesting. That is, if you do policy forecasts, you ought to have alternative simulations and not look at just one set of policy forecasts, which is more often than not the current policy. In part, my point of going clear back to five or six administrations ago to give an illustration of the difficulty of that was to give a historical perspective that the arguments on these issues aren't anything new, it's been with us forever. The other point that I think was implicit in what you said is one of the arguments that's always made against not making projections or forecasts. That is, people will take action based on your forecast and then your forecast will always be wrong. If that's true, I would say the forecaster is very successful even though when he or she evaluates those forecasts would show errors. But, if they have resulted in adjustments to policy or behavior, then I would think we have been very successful in alerting policy-makers or individuals to change their behavior based on something that's coming.

Dr. Kent: I would just add to that at the EIA we really do two types of forecasting. First, we do what I will call "policy-neutral" forecasting. That's the forecasting that you will find in all of our major periodic publications. Those forecasts are all based on the assumption that there will be no change in government policy, that what is today's policy will continue over the forecast period, whether it's 24 months or whether it's 30 years. That, of course, means that our forecast will be wrong because, quite obviously, policy is going to change. But the purpose of these policy-neutral projections is two-fold. First of all, we want to point out to policy-makers that this is probably where you're probably going to be if you don't change something. That becomes a very good basis, then, for policy decisions. The second reason that we do policy neutral forecasting is that it keeps us out of the political sphere of having to justify why did we assume that somebody was going to take thus and such an action, why did we assume that such a policy change was going to be made? And so most of our major forecasts then are policy-neutral to get around that situation. But a great deal of the forecasting we do — and I see Ron and Kay out here do a great deal of it and do it exceptionally well — is that we look at policy options and say, if you do this, this is what is going to happen; this is how the future is going to change; this is what the results are going to be. For example, if there is a carbon tax or if there are gasoline mileage standards or if relaxed licensing of nuclear power plants or passage of the Clean Air Act, what will be the result. You look at all of those policy options and say, if you make these changes in policy, then this is what the future is going to look like. That allows a more rational discussion and takes the burden off of us in trying to pick and choose among the policy options that might be out there.

Mr. Butz: Well, when you make population projections of the three major components — fertility, mortality, and immigration — one of those is very directly under the influence of policy. That's migration. The others are less directly influenced by policy. So when the Census Bureau people make those projections, they do in fact use alternative assumptions about the level of in-migration, which is closely tied to policy on immigration. Now, if the Census Bureau happened also to be in the business, let's say, of making cards for immigrants of some sort and we were giving a budget initiative, a budget request, to the Congress for printing those cards for five years from now, I think it would be very reasonable for the Congress or for the GAO to ask us to provide not a single projection of immigration into the United States, but several of them, and for somebody — we or you or the Congressional Research Service — to cost out the effects of those

different assumptions. So I think you have a good point. When a request is dependent in a fairly recognizable and quantifiable way on a legislative act or possible legislative act by Congress, it makes sense to me to ask the agency to submit at least a little bit of a sensitivity analysis of what difference it might make.

Question: I think in a certain sense we can summarize the discussion in that you want to do macro-coordination of federal forecasting. In a sense you're talking about coordination between your agencies. There is another aspect of this that might be very important also to all of you as administrators. That is, coordination of the forecasting out of your particular agency. In particular, the idea that you'd like to speak as much as you can with one voice to Congress and whoever your customers are in terms of the forecasts that you put out. So I was wondering if there could be a little bit of discussion on this kind of point, if possible, especially in terms of, say, expert computer systems that, I'll mention, coordination of what you're doing in terms of generating this consistency, of making the output of your agency more timely, more able to respond to your customers in a quick kind of way. That is, you can get your data and have your analysis done much more quicker. You know, you can pick out the outliers in the data easily. So you can end up using the high quality data and that sort of thing. If any of you have any comments on that, I'd like to hear them.

Mr. Fullerton: You spoke to the question of internal consistency of forecasts from a particular agency and also the use of expert systems to help achieve that, and asked the panel to comment.

Mr. Kutscher: Well, in the Bureau of Labor Statistics, if there is any blame for not coordination of a forecast, that's my fault because all the forecasts done by the Bureau of Labor Statistics are under me so it does not become a coordination issue rather than internal to my own staff. But I think you raised another issue somewhat tangentially, and that is, the coordination between the producers of the data that we use. In one sense, we are one of the largest users of the data produced in other parts of the Bureau of Labor Statistics. I didn't recite all of them but almost all of the employment series produced by BLS is projected by us. So the question is, how much coordination do I do with the group that produces the underlying employment data or the underlying productivity data or the underlying wage data or price data that I use in my model? There, we can perhaps be rightly brought to task because I'm not always sure that we have maximum coordination. I sometimes feel frustrated as an internal user of BLS data that I don't have as big a voice on the type of data collected, and when it's collected and published, as perhaps some outside users do. So I think one can always improve internal agency coordination in that way.

Dr. Kent: We're in exactly the same position. Anything that comes out under the EIA label has to have my approval and my sign-off on it. So, if we're not coordinated, it's because I don't know what I'm doing, which maybe explains why we're not very well coordinated. (Laughter.) But I think that's the only way that it can be successfully done. There has to be a single focal point through which all of your forecasts and all of your publications pass, and a central responsibility or your headquarters of your front office staff has to make sure that you're not producing dueling forecasts or coming up with data series which are incompatible with each other. Certainly, there's a lot of kicking within the EIA box. We don't all agree. In doing our major forecasts, that's an interoffice effort. We have what we call Category I review in which everybody else's work has to be signed off by every other major office indicating they agree with it. And it if can't pass Category I review, it doesn't go out. That way, we attempt to deal with these problems of making sure that we speak with one voice.

Mr. Butz: Well, I should pass on this but I won't. Within the Department of Commerce there are at least three organizational units that I know of that do population projections. One of them is in the Bureau of Economic Analysis and two of them are in the Census Bureau. And actually, those two are both under me. My impression is that the communication and coordination among the two, one of which does international population projections and one of which does U.S. population projections, could be better. And if that's the case, well then, that must be my fault.

Mr. Fullerton: We need to thank the panel for their thoughtful comments and to congratulate the audience for their provocative questions. We need also to release the panel. I remind you that our sessions start at 1:00 p.m. in various places. Consult the program and the map of Agriculture to find out where to go next.

Whereupon, the morning session of the conference was concluded.

Forecasting Practices in the Federal Government

Stuart Bretschneider, Associate Director of the Technology and Information Policy Program and Associate Professor of Public Administration, The Maxwell School, Syracuse University, Syracuse, NY 13244

Wilpen Gorr, Professor of Public Policy and Management Information Systems, School of Urban and Public Affairs, Carnegie Mellon University, Pittsburgh, PA 15213

Introduction

In recent years forecasting has become a major area of research in its own right. Since 1980 two professional societies and three journals have emerged devoted solely to forecasting[1]. The initial emphasis in forecasting research has tended to be methodological in nature, focusing on evaluation of alternative forecasting techniques. The evaluations have generally been based on "objective" measures of forecast performance, such as the mean absolute percentage forecast error, with very little concern for the "real" criteria at work in public and private organizations (Makridakis and Hibon 1979, Makridakis et al. 1982, Armstrong 1985). Though some studies have attempted to take into account organizational factors, little theoretical or empirical work has explicitly considered the context in which a forecast technique is used (Larkey and Smith 1984, Bretschneider and Gorr 1987, Bretschneider et al 1989, Kamlet et al 1989, Freenberg et al. 1989, Gentry 1989).

This paper thus seeks to develop a better understanding of forecasting by beginning with organizational context. Based on data collected from a mail survey of members of a group known as the Federal Forecasters, this paper explores organizational perspectives on forecasting in the federal government.

The Research Question

The principle focus of this paper is to begin to understand the complexity of forecast evaluation in organizations. Many criteria are at work in real organizations and those criteria reflect such diverse factors as mission, organizational culture, resource endowments, etc. Previous work in comparative organizational studies suggests that the environment of public and private organization may differ sufficiently to cause important differences in management (Bretschneider 1990). This implies that some factors affecting forecast evaluation may be unique to government or level of government.

Organizational size should have an effect on forecast criteria through resource dependency. Larger organizations are thought to have more slack resources which, if true, potentially down plays cost criteria and elevates alternative criteria in large organizations. Along with organizational size, the characteristics and endowments of the human resources assigned to forecasting also reflect the influence of resources. Here the background and experience of the forecaster should influence criteria. For example, experienced forecasters may place emphasis on continuity of forecasts, so as not to disrupt service delivery systems, and on conservative forecasting to avoid shortfalls of resources.

"Organizational environment" is determined by several factors. First, the position of a forecasting unit on an agency's organizational chart determines the purpose and level of forecasting. For example, if a forecasting group is closely linked to high levels of the agency, then presumably forecasting has some significance to policy making and control within the organization. In such cases the ability to evaluate alternative scenarios and contingencies is an important criterion for forecasting. Second, and particularly relevant to governmental forecasting, is the extent to which forecasting is influenced by political factors. Forecasting can be used as a policy tool to achieve desired ends, through consistent or selective biasing (Larkey and Smith 1984, Gentry 1989, Bretschneider et al 1990). In such cases forecasters must defend forecasts, even though forecasts are slanted. Third, the nature of a forecasting unit's mission is also affected by the extent to which it is intended to compete with other forecasts produced in the same or other agencies. Competition in forecasting tends to reduce intentional biases and may turn attention more to technical issues such as methods of estimation or the role of critical assumptions.

A third grouping of variables is the organizational management of forecasting; i.e., whether there are formal procedures to periodically evaluate forecast methods and processes. The literature on forecast techniques is vast but largely irrelevant to guide evaluations, since functioning organizations often make use of hybrid techniques that are pragmatic constructions (e.g., judgmental adjustment of econometric forecasts). Little is known about how the nature of an organization's forecast evaluation process affects forecast criteria. It is reasonable, though, to assume that organizations making use of formal evaluation processes might exhibit different criteria than those without such processes.

We can organize these proposed factors into a simple model (Figure 1). This is not meant to be more than an initial exploratory effort.

Data Collection. The sample frame for this study consisted of 259 members drawn from the 1989 and 1990 directories of a professional organization known as the Federal Forecasters. Members of the Federal Forecasters have either had forecasting as part of their job responsibilities in the federal government or have had an interest in forecasting. Association is voluntary, and the principle activity of the organization has been an annual conference where ideas are exchanged and discussed through the use of keynote speakers and presentation of papers.

As the above description implies, the use of the Federal Forecasters yields a sample frame of convenience, and cannot be construed to be a purely random sample of forecasters working in the federal government. Nevertheless, the sample does represent several federal agencies and has value for an exploratory study.

Survey Process. The survey process consisted of an alert letter, and an initial mailing of the survey with a cover letter two weeks later. There were two follow-up mailings for non-respondents. After the first follow-up mailing a 53% response rate was obtained and after the second the overall response rate was 60%.

Several types of responses were received other than completed surveys. These included individuals who returned uncompleted surveys because they no longer had job responsibilities which included forecasting, as well as individuals who referred to others in their organization who had already filled out a survey. Of the total number of respondents, over 75% are actively involved with developing or reviewing forecasts. In total 115 useable survey responses were obtained for a 45% sample.

Some individuals felt that the survey process was not sufficiently blind due to the use of control codes that identified individual respondents. This may have led to some problems of sample selection. On the other hand the pattern of responses by agency presented in Table 1 follows the distribution of membership, suggesting that at least agency was not a selection factor.

Responding Individuals and Organizations. Tables 2 through 5 describe the background of the individuals and organizations represented in the sample. In terms of educational background, 78% of the respondents have a Masters degree or higher and 47% have a background in economics. 75% of the respondents had at least five years of work experience in their current organization. In terms of experience in forecasting, 57% had taken a formal course in forecasting and the median respondent has eight years of experience in the forecasting area.

Finally the nature of the forecasting groups represented varied significantly. Not only were a wide spectrum of agencies represented in the sample (see Table 1) but these organizations varied significantly in terms of size. Though the mean group employed 13 full time equivalent staff members and had a budget of over $800,000, the median organization only employed five staff with a budget of $300,000. Clearly the sample contains a few very large forecasting units, while most employ 10 or fewer technical staff.

Forecasting Process and Forecast Evaluation. How do the member agencies of the Federal Forecasters make use of their forecasts? Table 6 shows that they use forecasts primarily for policy development and evaluation. The Federal Forecasters as a group tend not to make much use of forecasting for auditing or financial management. Of secondary interest are uses related to service delivery and the provision of public goods. Functions such as planning, program design and management of short term services demonstrated some value to respondents.

If the principle use for forecasts are policy analysis, who then are the ultimate users of the forecast information? Table 7 demonstrates that the heaviest users are agency technical staff. Interestingly the press is also viewed as a relatively major user for forecast information generated by the Federal Forecasters. On closer examination of the data, some of the responses reflect specific substantive areas, so that industrial groups are more likely to be rated higher by respondents from Energy, Commerce, Labor and Transportation, then from the military services, Treasury or Health and Human Services.

These results suggest that forecasting is a technical activity that serves other needs within the organization, though a fair amount of forecast information is used by political agencies and groups external to government.

Forecasting Process. Several questions regarding the forecasting process were also asked in the survey. There were questions with regard to the importance of different forecasting techniques, and the nature of evaluation processes. Table 8 summarizes responses on the level of use for a wide array of different forecasting methods. The dominant methodology in use is regression analysis and expert judgement. Note that the use of simultaneous equations, which require more sophisticated estimation are less often used.

Several different forms of time-series modeling were present in the list: time-regression, smoothing, and Box-Jenkins methods. Only the use of time-trend models with regression, however, were heavily used. These results are potentially explainable due to the large proportion of individuals having a formal background in economics. Over 46% of the respondents had economics as their single field of study, plus among the second largest grouping called "others", nearly half cited economics along with a second field of study. Regression analysis is the dominate estimation method in economics.

Also of interest is whether and to what extent these organization had formal evaluation processes for their forecasts. Table 9 summarizes responses to this question.

Most organization have some type of evaluation, though typically it is informal. Only 36% of the respondents have a formal forecast accuracy evaluation process.

Evaluation Criteria. The survey asked respondents to rate 25 different potential forecast evaluation criteria, the results of which are summarized in table 10. The single most important result from the analysis of criteria is that the pragmatic "explainability" criteria are far more important than any other type of evaluative criteria. By focusing on criteria such as defensibility of method and reasonableness of assumptions, the forecaster recognizes that the result must fit into a political process. The reliance on accurate data and linkages between data and forecasts reflect the professional values of the forecasters. Here technical forecasting relies on "good" data and "solid" theory to link data to forecasts, while fully recognizing that the result will ultimately be used within an adversarial political process.

Factors Effecting Criteria. At this point we return to the tentative model developed above and presented in Figure 1. The three major factors thought to influence valuation of criteria in use were 1) forecasting resources, 2) the forecast organizations environment, and 3) its management. In order to explore this model a series of operational measures for factors related to each of these areas was developed. Under forecasting resources the size of the forecasting organizations in full time equivalent (FTE) staff, the years of forecasting experience associated with the respondent and whether the respondent had taken a formal course in forecasting. While the experimental unit is the organization, the use of personal information about the preparer of the survey acts as a surrogate for the overall level of professionalism in the forecasting unit. High levels of experience and formal training in the respondent suggest similar traits among the other forecasting staff.

In terms of organizational environment, three general factors were used. The number of levels between the forecasting unit and the under-secretary level of the host agency should be inversely related to the relative importance of the unit to the organizational mission. The organizational distance between forecasting and the under-secretary might also reflect the extent to which the forecasting unit is exposed to political pressure. Two questions on the survey attempted to focus on the extent of political pressure faced by the forecasting unit: 1) the extent to which political pressure is used to adjust forecast to meet political ends and 2) the extent to which political pressure make forecasts subject to review and audit by outside groups. Both of these operational measures used five point Likert scales. A final variable used to characterize the organizational environment of federal forecasting was a binary variable used to indicate the presence of absence of another group forecasting the same phenomenon. Here our concern is that the existence of a competitive forecast could create a qualitatively distinct environment for forecasting (Bretschneider and Gorr 1987, Bretschneider et al 1989).

The last area of influence is the actual forecasting practices. Though methods and techniques could have an influence, the large number of alternative in use made this a poor operational measure. Of possibly greater importance, though, was whether the forecasting unit had any type of formal forecast accuracy evaluation. Consequently a binary variable was used to distinguish between organizations with no or only informal evaluation and those with a formal evaluation process.

Ordinary least squares (OLS) regression was used to investigate the extent to which these factors influence the valuation of forecast evaluation criteria by survey respondents. For the full model, Table 11 presents those regression coefficients that were statistically significant at 10% or less plus the adjusted R-Square and sample size for each of the 25 criteria. The criteria have been grouped into categories as they were in the original survey instrument.

In terms of overall explanatory power the proposed model has only modest power. For a few criteria, such as the reasonableness of assumptions, accuracy, bias, and adjusta 0bility of forecasts, the estimated models produce reasonable results. Some of the descriptive variables act in a consistent way in influencing the valuation process. For example, technical and quantifiable criteria are enhanced by having had a formal course in forecasting. This was true for accuracy of data, forecast accuracy, serial correlation, simple benchmarking, adjusting i nputs and standardized methods. Not surprisingly, most of these topics are handled in formal courses on forecasting, particularly from an econometric perspective (Pindyck and Rubinfeld 1991).

The presence of another agency forecast or competitor also produces consistent results, where reasonableness of assumptions, adjustability of forecasts, and policy evaluation criteria are all enhanced. With competition, justification for a forecast becomes more important, hence the importance of reasonable assumptions. Finally political pressure to adjust forecasts to political ends consistently decreased the value of criteria, though only in two cases was this significant. Several of the descriptive factors generated both positive and negative valuation effects, but to some extent are still consistent.

The reversal of effects can in part be explained by thinking of criteria as either technical/scientific versus practical/political. For example decreasing the presence of serial correlation in a forecast is highly technical, so having had a formal course in forecasting and the presence of a formal evaluation process tended increase the importance of this criterion. On the other hand, increased experience and political pressure within the same model tended to reduce its value. Experience in forecasting tends to enhance practical criteria like defensibility of methods, manpower costs, and multiple uses while decreasing technical criteria. Similarly if a formal evaluation process is used, technical criteria increase in importance while practical political criteria decrease. Political pressure to external review behaves in a somewhat similar fashion. Increased sophistication of methods is decreased while focus on bias, simple benchmarking and standardized reporting are enhanced.

The final two descriptive factors are the size of the forecast organization and its importance as measured by the number of reporting levels between the forecasting unit and the under-secretary of the agency. Neither of these two variable operate in a consistent or predictable fashion. In fact the levels variable operated inversely to what might be expected. As the forecasting function becomes more importance, pragmatic criteria should become more valuable since the forecasts are becoming more important to the policy process. Yet here, the results are that technical criteria, particularly sophistication of techniques become more important while the ability to link forecast to data is decrease in importance.

These two variables also contained several missing responses in otherwise completed surveys. It is also possible that the highly skewed nature of the FTE size variable could be influencing the results. By re-estimating the model without these two terms the sample increased by 12 to 15%. The results of this estimation are presented in table twelve. The results are very similar to those reported in table eleven though the general effects of the remaining coefficient is more consistent across evaluation criteria.

Conclusions. This paper has begun to investigate how organizational environment and context affect the process of forecasting, and, in particular, the relative importance of various evaluative criteria for assessing forecasts. Using a sample of forecasting organizations in the federal government, we find some environmental factors seem to influence the relative importance various criteria play in the evaluation process. For example the presence of a competing agency or group generating a forecast, political pressure, the presence of a formal evaluation process, and the general level of professionalism and experience in forecasting all tend to influence relative importance of criteria.

Interestingly the most technical and objective criteria such as bias, accuracy, and serial correlation can be relatively well explained. In particular the more pressure for external review, the presence of a formal evaluation process, and more professionalism in the staff (e.g. having had a formal course in forecast evaluation) the more important these criteria become. Negative value for these technical criteria came from work experience and political pressure to adjust forecast. What is striking about these results is that they suggest that professionalism and training tend to emphasize technical measurable criteria but fail to reflect on the important of feasibility and the need for forecasts which can be flexible enough to match the political factors faced by forecasters.

It must be remembered that the sample analyzed here is a select group that is not generally representative of the federal government; e.g., many important forecasting groups such as federal reserve banks are not well represented. The sample also tends to over represent policy-oriented uses and is heavily dependent on econometric/regression techniques. Yet some of these unique characteristics are important. In particular, reliance on purely technical criteria such as accuracy and bias, especially for this group, seems ill advised.

To illustrate, consider a recent forecast audit compiled by the GAO which found fixed bias in such forecasts and that certain simple non-regression based time-series extrapolation benchmark techniques were superior in terms of bias (GAO 1991). This might suggest an alternative approach to forecasting, yet that conclusion fails to recognize to complex organizational environment faced by these organizations. It is impractical to substitute an extrapolative method for one based on theory, when evaluation of scenarios and evaluation of alternative policies is a major reason for forecasting. This leaves the professional forecaster faced with some difficult choices: does she or he rely on technical criteria with its safeguards of professionalism or does a more balanced set of criteria need to be considered which include issues of political feasibility, and manipulability. The empirically derived answer seems to be one of a mixture of criteria, but a mix that aims to integrate "explainability" and "manipulability" along with policy utility and the more technical "objective" criteria.

References

Armstrong, J. Scott, ***Long Range Forecasting: From Crystal Ball to Computer***, Second Edition, New York: John Wiley and Sons, 1985.

Bretschneider, S. and W. Gorr, "State and Local Government Revenue Forecasting," in Makridakis, S. and Wheelwright, S. eds., ***The Handbook of Forecasting***, 2nd Edition, New York: John Wiley and Sons, 1987.

Bretschneider, S., W. Gorr, G. Grizzle, and E. Klay, "Political and Organizational Influence on Forecasting Accuracy: Forecasting Sales Tax Receipts," ***International Journal of Forecasting***, Vol. 5, No. 3 (1989), pp. 307-319.

Bretschnedier, S., "Management Information Systems in Public and Private Organizations: An Empirical Test," ***Public Administration Review***, Vol. 50, No. 5 (1990) pp. 536-545.

Cassidy, G., M. Kamlet, and D. Nagin, "An Empirical Examination of Bias in Revenue Forecasts by State Government," ***International Journal of Forecasting***, Vol. 5, No. 3 (1989), pp. 321-333.

Feenberg, D.R., W.M. Gentry, D. Gilroy, and H.S. Rosen, "Testing the Rationality of State Revenue Forecasts," ***Review of Economics and Statistics***, Vol. 71, No. 2 (1989), pp. 300-308.

GAO Report, ***Short-Term Forecasting: Accuracy of USDA's Meat Forecasts and Estimates***, GAO/PMED-91-16, May, 1991.

Gentry, W.M., "Do State Revenue Forecasters Utilize Available Information?", Unpublished Manuscript, Princeton University, 1989.

Kamlet, M.S., D.C. Mowery, and T. Su, "Whom Do You Trust? An Analysis of Executive and Congressional Economic Forecasts," ***Journal of Policy Analysis and Management***, Vol. 6, No. 3 (1987), pp. 365-384.

Larkey, P., and R. Smith, "The Misrepresentation of Information in Government Budgeting," in L. Sproull and P. Larkey (eds.) ***Advances in Information Processing in Organizations***, Vol. 1, Greenwich, Conn.: JAI Press, 1984.

Makridakis, S. and M. Hibon, "Accuracy of Forecasting: An Empirical Investigation," ***Journal of the Royal Statistical Society***, Series A, Vol. 142, part 2, (1979), pp. 97-145.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods, ***Journal of Forecasting***, Vol. 1 No. 2 (1982), pp. 111-154.

Pindyck, R. and D. Rubinfeld, ***Econometric Models and Economic Forecasts***, 3rd Edition, New York: McGraw Hill, 1991.

Footnotes

[1]. In 1981 the International Institute of Forecasters (IIF) was formed and began publication of the ***Journal of Forecasting*** (JoF). In 1985 the IIF shifted its sponsorship to the ***International Journal of Forecasting*** (IJF), though the JoF is still published separately by John Wiley and Son. Also during this decade a professional group of business forecasters was formed which continues to publish ***The Journal of Business Forecasting***.

Figure 1. Model of Forecast Evaluation Criteria.

Forecasting Resources	Forecasting Organizational Environment	Forecasting Organizational Management
a. Slack	a. Organization Chart	a. Techniques
b. Human	b. Political Influence	b. Process
	c. Competition	

Relative Importance of Forecast Evaluation Criteria

Table 1. Distribution of Respondents by Agency

	Sample Percent	Population Percent
Agriculture	15	15
Health and Human Services	14	14
Commerce	12	10
Labor	11	11
Army	5	4
Energy	5	5
General Accounting Office	5	2
Navy	5	6
Transportation	5	5
Treasury	5	3
Veteran Affairs	4	5
Interior	2	2
Others	2	3
Tennessee Valley Authority	2	2
Governors, Federal Reserve	1	1
Congressional Budget Office	1	2
Coast Guard	1	1
Department of Defense	1	1
Office of the President	1	1
Farm Credit Administration	1	1
Housing & Urban Development	1	2
National Science Foundation	1	1
Office of Personnel	1	1
Postal Service Commission	1	1

Table 2. Educational Background of Respondents

a) Highest Degree

	Frequency	Percent
Associate Degree	1	0.9
Bachelor Degree	23	20.4
Master Degree	49	43.4
Doctorate	37	32.7
Other	2	1.8

b) Area of Study

	Frequency	Percent
Economics	52	46.8
Others	26	23.4
Statistics/Math	13	11.7
Engineering	5	4.5
Business Adminis	5	4.5
Public Administr	5	4.5
Sociology	3	2.7
Anthropology	1	0.9
Chemistry	1	0.9

Table 3. Work Experience and Organizational Tenure for Respondents

a) Years worked in current organization

N	113
Mean	11.34
Std Dev	8.60
Maximum	41
Third Quartile	15
Median	9
First Quartile	5
Minimum	1

b) Years worked in current position

N	113
Mean	6.15
Std Dev	4.91
Maximum	21
Third Quartile	10
Median	4
First Quartile	2
Minimum	1

Table 4. Respondents Background in Forecasting

a) Formal course on forecasting

	Frequency	Percent
No	48	42.9
Yes	64	57.1

b) Years worked in forecasting area

N	110
Mean	11.01
Std Dev	9.10
Maximum	35
Third Quartile	18
Median	8
First Quartile	3
Minimum	0

Table 5. Organizational Size for Respondents

a) Size in Full Time Equivalent (FTE) forecasting positions

N	104
Mean	13.15
Std Dev	19.37
Maximum	100
Third Quartile	12.5
Median	5
First Quartile	2.5
Minimum	0

b) Size in forecasting budget

N	73
Mean	$880,767
Std Dev	$1,623,473
Maximum	$10,000,000
Third Quartile	$900,000
Median	$300,000
First Quartile	$100,000
Minimum	$0

Table 6. Importance of Different Uses or Purposes for Forecasts.
Scale: NOT IMPORTANT = 1 TO ESSENTIAL = 5
ZERO INDICATED NOT APPROPRIATE

Purpose of Forecast	N	Mean	Median
Proposed policy change	114	3.3	4
Implemented policy change	114	3.2	4
Strategic planning	114	2.5	3
Program design	114	2.2	3
Short-run services demand	114	2.1	2
Capacity planning	114	2.1	2
Budget planning & design	114	1.9	1
Budget gap/pol. planning	114	1.6	0
Budget monitoring	114	1.3	0
Capital budgeting	114	1.1	0
Audits/Federal agencies	114	0.9	0
Audits/State&Local govern	114	0.7	0
Audits/other Fed. branches	114	0.7	0
Fund balance investment	114	0.6	0
Audits/regulated industry	114	0.6	0

Table 7. Forecast Users
Scale: NO USE = 1 TO HEAVY USE = 5

Forecast Users	N	Mean	Median
Agency technical staff	109	3.229	3
Agency management	107	3.103	3
Other external groups	64	2.766	3
Other agency tech. staff	107	2.710	3
Press	105	2.600	3
Senate members/staff	110	2.436	2
House members/staff	109	2.431	2
Industrial groups	106	2.396	2
Other agency management	101	2.337	2
OMB	103	2.117	2
Judicial organizations	97	1.320	1

Table 8. Forecasting Methods Used
Scale: VERY RARE = 1 TO ALMOST ALWAYS = 5
ZERO INDICATED NOT APPROPRIATE

Forecasting Methods Used	N	Mean	Median
Time-regression extrapolation	114	3.2	4
Experts for independent variables	114	3.1	4
Aggregate regression equations	114	2.9	3
Experts for dependent variables	114	2.9	3
Judgmental & quantitative	114	2.8	3
Adjustment for structural chng	114	2.7	3
Multiple quantitative forecasts	114	2.7	3
Disaggregated regression equation	114	2.6	3
'Eyeball' extrapolation of data	114	2.5	3
Smoothing for extrapolation	114	2.3	3
Simult. equation econometric	114	2.3	2
Multiple judgmental forecasts	114	2.2	2
Box-Jenkins method	114	1.6	2
Intentions survey	114	1.4	1

Table 9. Nature of Forecast Accuracy Evaluation Process

	Frequency	Percent
No forecast evaluation	12	10.9
Informal irregular evaluation	32	29.1
Informal regular evaluation	26	23.6
Formal irregular evaluation	13	11.8
Formal regular evaluation	27	24.5

Table 10 Importance of Different Forecast Evaluation Criteria
Scale: NOT IMPORTANT = 1 TO ESSENTIAL = 7

Forecast Evaluation Criteria	N	Mean	Median
Reasonable assumptions	106	5.7	6
Sensible & accurate data	107	5.5	6
Output forecasts link to data	105	5.5	6
Defendable methods	106	5.4	6
Ability to evaluate scenarios	108	4.7	5
Ability to adjust forecasts	107	4.7	5
Policy alternative evaluation	108	4.5	5
Ability to ID structural change	107	4.5	5
Ability to adjust inputs	107	4.4	4
Use of standard data	108	4.4	4
Multiple uses for forecasts	109	4.3	4
Accuracy (Mean Squared Error)	100	4.3	4
Ability to agg & disagg forcasts	108	4.3	4
Coordination of timing	108	4.2	4
Systematic bias (Mean Error)	100	4.0	4
Ability to present range of vals	108	4.0	4
Ability to identify outliers	107	3.9	4
Use of standard methods	107	3.8	4
Standard forecast reporting use	107	3.8	4
Manpower costs	109	3.7	4
Ex. of serial correlation	98	3.3	3
Simple benchmarking	105	3.3	3
Sophistication of technique	100	3.3	3
Computer costs	108	3.1	3
Multiple benchmarking	104	3.1	3

Table 11. Factors Effecting Importance of Forecast Evaluation Criteria
(regression coefficients are significant at 10% or less and
values in parenthesis are average rank order for importance)

Dependent Variable	FTE	Political Pressure Levels	Political Pressure Other Agency	Adjust Forecast	Review/ Audit	Formal Evaluation	Forecast Course	Experience	Adjusted R-Squared	N
EXPLAINABILITY (3)										
Defensible Method(4)	-	-	-	-	-	-	-	0.0324	0.0733	87
Reasonable Assum(1)	-	-	0.6527	-	-	0.4860	-	-	0.1189	87
Accurate Data(2)	-	-	-	-	-	-	0.4777	-	0.0463	88
Forecast Link Data(3)	-	0.1368	-	-	-	-	-	-	0.0104	86
TECHNICAL (18)										
Bias(15)	-	-	-	-	0.2178	-	-	-	0.0963	83
Accuracy(12)	-	-	-	-	-	1.4592	0.7454	-	0.1969	83
Serial Correlation(21)	-	-0.0061	-	-0.3641	-	0.6403	0.8647	-0.0568	0.2217	81
Sophistication(23)	-	-0.2198	-	-	-0.3368	-	-	-	0.0570	83
COMPARATIVE (24)										
Simple Bench(22)	-	-	-	-	0.3309	-	0.5689	-	0.0984	85
Multiple Bench(25)	-	-	-	-	-	-	-	-	0.0000	85
MANAGING UNCERTAINTY (11)										
Range of Values(16)	0.0237	-	-	-	-	-	-	-	0.0724	89
Eval. Scenarios(5)	0.0295	-	-	-	-	-0.8427	-	-	0.0536	89
Identify Outliers(17)	0.0319	-	-	-	-	-	-	-	0.0595	88
Structure Change(8)	0.0259	-	-	-	-	-	-	-	0.0649	88
MANIPULABILITY (9)										
Aggregate/Disagg(13)	-	-	-	-	-	-	-	-	0.0000	89
Adjust Inputs(9)	-	-	-	-	0.2867	-	-	-	0.0421	88
Adjust Forecasts(6)	-	-	0.8612	-	-	0.9316	0.5650	-	0.1316	88
COSTS (22)										
Manpower Cost(20)	-	-	-	-	-	-	-	0.0496	0.0000	90
Computer Cost(24)	-	-	-	-	-	-	-	-	0.0000	89
COORDINATION (16)										
Standardize Data(10)	-	-	-	-	-	-	-	-	0.0000	89
Standardize Meth(18)	-	-	-	-	-	-	-	-	0.0000	88
Standardize Rept(19)	-	-	-	-	0.2751	-	0.7081	-	0.0325	88
POLICY UTILITY (11)										
Timing(14)	-	-	-	-	-	-	-	-	0.0000	89
Multiple Uses(11)	-	-	-	-	-0.3360	-	-	0.0307	0.0614	90
Policy Evaluation(7)	-0.0172	-	0.5854	-	-	-	-	-	0.0560	89

Table 12. Factors Effecting Importance of Forecast Evaluation Criteria
(regression coefficients are significant at 10% or less)

Dependent Variable	Political Pressure: Other Agency	Political Pressure: Adjust Forecast	Review/ Audit	Formal Evaluation	Forecast Course	Experience	Adjusted R-Squared	N
EXPLAINABILITY (3)								
Defensible Method(4)	-	-	0.1855	-	-	-	0.0520	101
Reasonable Assump(1)	0.5566	-	-	0.4249	-	-	0.0690	101
Accurate Data(2)	-	-	0.1531	-	0.3938	-	0.0353	102
Forecast Link Data(3)	-	-	-	-	-	-	0.0000	99
TECHNICAL (18)								
Bias(15)	-	-	0.2619	0.5113	0.7360	-	0.1428	95
Accuracy(12)	-	-	-	1.3800	0.7667	-	0.2074	95
Serial Correlation(21)	-	-0.4878	0.2183	0.6033	0.8765	-0.0443	0.2160	94
Sophistication(23)	-	-0.3990	0.2519	-	-	-	0.0315	96
COMPARATIVE (24)								
Simple Benchmark(22)	-	-0.2782	0.3919	-	0.5689	-	0.1443	99
Multiple Bench(25)	-	-	-	-	-	-	0.0000	98
MANAGING UNCERTAINTY (11)								
Range of Values(16)	-	-	-	-	-	-	0.0000	102
Eval. Scenarios(5)	-	-	-	-	-	-	0.0000	102
Identify Outliers(17)	-	-	-	-	-	-	0.0000	101
Structure Change(8)	-	-	-	-	-	-	0.0000	101
MANIPULABILITY (9)								
Aggregate/Disagg(13)	-	-	-	-	-	-	0.0000	102
Adjust Inputs(9)	-	-	0.3147	-	-	-	0.0822	101
Adjust Forecasts(6)	0.9524	-	-	0.7831	0.5496	-	0.1223	101
COSTS (22)								
Manpower Cost(20)	-	-	-	-	-	0.0370	0.0000	103
Computer Cost(24)	-	-	-	-	-	-	0.0000	102
COORDINATION (16)								
Standardize Data(10)	-	-	-	-	-	-	0.0000	102
Standardize Meth(18)	-	-	-	-	-	-	0.0000	101
Standardize Report(19)	-	-	0.2742	-	-	-	0.0252	101
POLICY UTILITY (11)								
Timing(14)	-	-	-	-	-	-	0.0000	102
Multiple Uses(11)	-	-0.3611	-	-	-	-	0.0534	103
Policy Evaluation(7)	0.5978	-	-	-	-	-	0.0411	102

U.S. Department of Agriculture Forecasts: Personal Comments on GAO's Forecast Evaluation Methodology

David R. Solenberger[1], U.S. General Accounting Office

The U.S. General Accounting Office (GAO) increasingly is being asked to answer questions about the future. Forecast evaluation of historic forecast error supports our capacity to answer these questions well. This is because systematic methods for dealing with questions about the future can be more efficient and yield sounder, better documented answers than more informal methods do.

Many methods exist to deal with forward-looking, future-oriented questions. Collectively, they are referred to as "prospective methods" to distinguish them from "retrospective methods" which are approaches designed to answer questions about what is happening now or what has happened in the past.

This paper is based on David Solenberger, Fred Light, and Gary Billen's work on USDA forecast evaluations. Stuart Breschneider, an associate professor at Syracuse University, worked as a consultant for the GAO evaluations.

The appropriateness of a forecast methodology can be evaluated in two ways. One is to verify that the methods accurately reflect the relationships of such factors as market prices, supply, and demand. The second is to evaluate the forecast results by measuring historical accuracy and by comparing that accuracy to results from other methods. In this forecast evaluation paper, we concentrate on forecast accuracy measures, and not methods.

Seven specific evaluation questions are used in assessing forecast accuracy. The methodology emphasizes comparing the forecasts to the actual subsequent event. The evaluation questions include the following:

1. What methodology is used for forecasting the event?
2. Who uses the forecasts?
3. How can forecast accuracy be measured?
4. How accurate are the forecasts?
5. Are the errors we identified "reasonable?"
6. What are the implications of forecast error on private sector as well as on the government policy, program, and budget decisions?
7. How can forecasts be improved?

This paper offers comments on the methods used to answer the evaluation questions. Appendix I shows an extensive literature search. Appendix II identifies selected GAO reports which address forecasting issues.

The reader may wish to present and address the evaluation questions in the order presented in this paper. Other options are certainly possible. However, we strongly suggest first determining that the variables being evaluated are important enough to warrant the time needed for a full evaluation.

Forecasting Methodology. Methods for forecasting can be classified into three different types; (1) modeling such as input-output and econometric, (2) trend extrapolation, and (3) judgmental and/or consensus.

- Modeling is the representation of a system and its elements or variables and the relationship between the elements that govern their interaction. The representation may be theoretical, mathematical, physical or a combination of these. Types of modeling can involve econometric or input-output analysis.
- Trend extrapolation or regression involves projecting historic trends into the future. These projections involve one or more variables.
- Judgmental forecasts involve no specific analytical technique, but instead opinion. The field of psychology developed two types of judgmental forecasts; delphi and panel-of-experts. Those offering the judgments about future events are generally considered experts about the subject matter being forecast. Consensus forecasts do have structured techniques for arriving at a decision. Consensus, or combined forecasts, represent the construction of a forecast based on several forecasts. These forecasts may be mechanically combined in a number of ways.[2]

For the purpose of forecast evaluation, we do not evaluate the methodologies used for forecasting. Instead, we concentrate on evaluating the forecast accuracy. Descriptive information may be useful in addressing the other evaluation questions.

Forecast Users. Forecast users generally fall into three types: (1) private sector decisionmakers involving production, marketing, and/or speculation; (2) policy-making and analysis, and (3) program implementation. Private sector production and/or marketing decisions can involve commodity planting or production decisions. Private sector speculation is traditionally related to commodity futures markets. Policy-making involves legislation, budgeting, and international trade negotiations. Program implementation includes agency decision making where forecasts are required such as for justifying meat purchases, offering insurance, assessing the credit worthiness of loans, modifying budgets, and assessing general economic conditions.

Forecast Accuracy Measures. Forecast accuracy measures rely on the concept of errors. Error (E) is measured as the actual (A) value less the forecast (F); that is, $E = A - F$. Single forecast errors may be positive or negative. Single errors, however, do not have much value for gauging the quality of the forecast. But multiple forecasts made over varied times can be used to show how accurately a forecasting procedure is working. Calculated in this way negative errors are overestimates whereas, positive errors are underestimates.

The sum of the two components of forecast error—random and bias error—is "total error." To analyze forecasting methods, the single forecast error can be separated into two parts. One part is called "random error" and it varies unsystematically from one forecast to the next. The other part is called "bias error" and it remains constant for any particular forecasting procedure. Total error is measured with absolute measures (that is, negative and positive signs are not considered). Measurement of the random and bias error components, however, involves consideration of the negative and positive signs of single errors over time. These two partially offset each other, thus canceling out random error which is unavoidable and identifying bias error which can be reduced. Research has shown that the causes of bias error can frequently be isolated and corrected. Bias error can result from many factors including problems of design, methodology, measurement instruments, input data, or subjectivity (conscious or unconscious) on the part of the analyst.

In analyzing error in multiple forecasts, we concentrated on total (absolute) error and bias error measures. We display error rates first using graphical individual percent error rates, then

summary error rates. A number of summary error measures exist. Measures can either be in values or percentages. Percentage formulas we used for bias error measures include mean percent error (MPE), trimmed mean percent error (TMPE), and weighted mean percent error (WMPE). Total error measures include mean absolute percentage error (MAPE), adjusted mean absolute percentage error (AMAPE), and root mean square percentage error (RMSPE). Even more complex formulas are available if needed.

We found the simplest measures to be preferable, unless other measures exhibit significantly different results. For example, in identifying high error rate years, we found all summary error measures would identify the same year as the highest, even though the relative ranking of forecasts in some individual years may differ. When measures show large error differences by year, the measure used becomes more important.

Observed Forecast Error Rates. We can display forecast error rates using graphical or tabular techniques. Again, the issue is providing information to the forecast user in as simple and easy-to-use manner as possible.

Our experience indicates that summary forecast error rates exhibit several of the following traits;

- Error increases as the forecast period lengthens.
- Total error generally exceeds that of bias error, However, if total error and bias error are equal, that is, all of the errors are in the same direction, then a serious problem exists.
- Bias error may not approach 0 over a multi-year period. Even when it does, bias error may exhibit a cycle where error is consistently overestimated in some years and then consistently underestimated in others.

We may also wish to determine why the forecast errors occur. Forecast evaluation experts claim that, when forecasters systematically evaluate their past errors improvements in future forecast error can occur. Whether a forecast evaluation study should devote the resources to evaluating why errors occur is debatable. We found many of the errors occur as a result of reasons that forecasters should have considered. This is not to say, however, that the forecasters should not assess their errors. In all cases, forecasters should routinely assess and document why their forecasts are in error.

Error Reasonableness. Because forecasting is based on incomplete knowledge about the future, that some level of error will occur should be expected. Thus, total and bias error measures by themselves do not provide a basis for evaluating what level of error in forecasts is "reasonable." To determine this, it is necessary to compare the errors to other available benchmarks as a way of determining whether smaller error rates are possible. "Reasonable" would imply that both total and bias error are small and that no better forecasts are readily available.

A benchmark forecast is another forecast for the same variable that can be used for comparison purposes. Competitive and naive are two types of benchmarks often used. Competitive forecasts are those made by other individuals or groups. Naive forecasts use historical information and simplified models or averages of recent historic periods. Benchmark forecasts should be simple, low-cost alternatives.

Benchmarks demonstrate that lower bias error forecasts with similar total error rates for these years are possible. We found that the benchmark forecast results may show less bias error with similar total error. While benchmark forecasts cannot replace a basic forecasting method, they can be helpful in identifying where improvements are needed. Benchmarks are especially valuable for long term forecasting.

Once different error rates are calculated, some assessment of whether the differences are significant is in order. Statistical measures are available to facilitate this assessment.

Implications of Forecast Error On Policy, Program, and Budget Decisions. Even when forecast errors exist, and we demonstrate that they are not reasonable, we must determine whether those "unreasonable" errors resulted in an adverse impact. We can determine adverse impact through modeling and sensitivity analysis, obtaining comments from program experts, or applying Armstrong's[3] methodology. This is the ultimate test of "significance." Early in the evaluation, we should determine that the variables being analyzed are important.

In our long term forecasting report, we used USDA's policy simulation models to evaluate potential budgetary impacts.[4] We found that if the historic bias error rate exhibited in the baseline forecasts for crop years 1981-88 continued and if the farm program provisions in place at the start of 1990 had been extended for 5 more years, the $47.1 billion outlay estimate used in the administration's January 1990 budget submission might have been underestimated by $19.5 billion.

Suggestions for Improving Forecast Accuracy. We believe that properly managing and evaluating the forecasting process will result in more accurate forecasts. A prior GAO report recommended specific improvements for USDA forecasts, which was termed "the process of forecast management."[5] The Food Agricultural, Conservation, and Trade (FACT) Act of 1990 used these recommendations to suggest that the Secretary of Agriculture designate a single organization to manage its commodity program forecasting and establish a quality control program to—(1) systematically identify the source of forecasting errors, (2) maintain records of data used for supply and utilization forecasts, (3) document its forecasting methods, and (4) correct weaknesses in its various forecasting components.

Summing It All Up. Our forecast evaluation methodology measures error, assesses whether that error is reasonable, and then determines whether that error has a significant operational impact. After determining that the forecast error has significant impacts, we then offer suggestions for reducing error rates.

Although this methodology has been used in a number of GAO evaluations, each forecast evaluation differs. For example, our meat evaluation emphasized any error, since both low and high price forecasts affect meat producers. The federal budget impact study, however, mostly relates to high bias error for price and export forecasts. Current FCIC and EPA superfund forecast evaluation addresses what to do when no actuals are available.

Footnotes

1 Mr. Solenberger has worked at GAO's Kansas City Regional Office for 21 years. During the last 10 years he served as the regional operations research analyst and now as a project manager for program evaluations. This paper reflects Mr. Solenberger's personal opinions, and not necessarily those of the US General Accounting Office. The attached bibliography refers to specific GAO reports which address forecast evaluation issues.

[2] A considerable amount of literature exists concerning combined forecasts. For references to over 200 articles, see Robert T. Clemen's article, "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting*, Vol. 5, 1989, pages 559-583.

[3] Dr. Scott Armstrong. *Long Term Forecasting; From Crystal Ball to Computer*, John Wiley & Sons, New York, Appendix A, pages 452-458, 1985.

[4] *USDA Commodity Forecasts; Inaccuracies Found May Lead to Underestimates of Budget Outlays*, PEMD-91-24, August 1991, pages 39-40.

[5] U.S. General Accounting Office, *USDA's Commodity Program: The Accuracy of Budget Forecasts*, GAO/PEMD 88-8 (Washington, D.C.: April 1988), p. 75-76.

Bibliography

Armstrong, J. Scott, *Long Range Forecasting: From Crystal Ball to Computer*, 2nd ed., New York: John Wiley and Sons, 1985.

Ascher, William, *Forecasting: An Appraisal for Policymakers and Planners,* Baltimore: Johns Hopkins University Press, 1978.

Becker, Geoffrey S. *CRS Report for Congress: Provisions of the 1990 Farm Bill*, Congressional Research Service, 90-553 ENR, (Washington, D.C.), November 16, 1990.

Becker, Geoffrey S. *CRS Issue Brief: Agriculture and the Budget*, Congressional Research Service, IB87032, (Washington, D.C.), Updated December 31, 1990.

Bessler, David A., and Jon A. Brandt. "An Analysis of Forecasts of Livestock Prices." Texas Agricultural Experiment Station paper number 22437, May 1990.

Bretschneider, Stuart, *Forecasting: Some New Realities*, Metropolitan Studies Program, Syracuse University, Syracuse, New York, Dec. 1985.

Bullock, J. Bruce, "Social Costs Caused by Errors In Agricultural Production Forecasts, *American Journal of Agricultural Economics,* Feb. 1976, 76-80.

Clemen, Robert T. "Combining Forecasts: A Review and Annotated Bibliography," for references to over 200 articles. *International Journal of Forecasting*, Vol. 5, 1989, Pages 559-583.

Elam, Emmett, and Bruce L. Dixon, "Examining the Validity of a Test of Futures Market Efficiency." *The Journal of Futures Markets*, 8:3 (1988), 365-72.

Ferris, John. *Evaluation of Forecasts from the Annual AAEA Outlook Survey.* Presented at the annual meeting of the American Agricultural Economics Association; Reno, Nevada; July 1986.

Fox, Karl A., *An Appraisal of Deficiencies in Food Price Forecasting for 1973, and Recommendations for Improvement*, Washington, D.C., Nov. 29, 1973.

Garcia, Philip, et al. "Pricing Efficiency in the Live Cattle Futures Market: Further Interpretation and Measurement." *American Journal of Agricultural Economics*, February 1988, pp. 162-69.

Garcia, Philip, Michael A. Hudson, and Mark L. Waller. "The Pricing Efficiency of Agricultural Futures Markets: An Analysis of Previous Research Results." *Southern Journal of Agricultural Economics*, 20:1 (July 1988), 119-30.

Gorr, W. L., "Use of Special Event Data in Government Information Systems," *Public Administration Review,* 46 (Nov. 1986), 532-39.

Gunnelson, G., W.D. Dobson, and S. Pamperin, "Analysis of the Accuracy of USDA Crop Forecasts," *American Journal of Agricultural Economics,* Nov. 1972, p. 639-645.

Haidacher, Richard C., and Matthews, Jim L., *Review of Forecasting in the Economic Research Service*, Washington, D.C.: January 1977

Just, Richard E., "The Impact of Less Data on the Agricultural Economy and Society," *American Journal of Agricultural Economics*, December 1983, 872-881.

Just, Richard E. and Rausser, Gordon C., "Commodity Price Forecasting With Large-Scale Econometric Models and the Futures Market," *American Journal of Agricultural Economics,* May 1981, 197-208.

Kramer, Carol S., *The Political Economy of U.S. Agriculture,* "Agriculture and the Failure of the Budget Process," Charles H. Reimenschneider and Robert E. Young II, Resources for the Future, 1989.

Makridakis, Spyros, et al., *The Forecasting Accuracy of Major Time Series Methods,* New York: John Wiley and Sons, 1984.

Matthews, Jim L., "Review and Appraisal of Forecasting Activities in the Economic Research Service," Initial Report to the ERS Administrator, Washington, D.C.: May 1974.

Milonas, Nikolaos T., "The Effects of USDA Crop Announcements on Commodity Prices, " *Journal of Futures Markets,* October 1987, pp. 571-589.

Strommen, Norton D., "A Global Perspective on Weather Trends—The Most Unpredictable and Least Controllable Nature Resource," Presented at the World Food Production Symposium, November 6, 1989 in Rio De Janeiro, Brazil.

Sumner, Daniel A., and Rolf A. E. Mueller. "Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions." *American Journal of Agricultural Economics*, February 1989, pp. 1-8.

U.S. Congress. *Review of the Activities of the US Department of Agriculture's Crop Reporting Board*, Hearing before the Subcommittee on Department Operations, Research, and Foreign Agriculture of the Committee on Agriculture, House of Representatives, Serial No. 98- 61, Washington, D.C.: U.S. Government Printing Office, February 23, 1984.

U.S. Congress. *The Outlook for Farm Commodity Program Spending Fiscal Years 1990-1995,* Washington, D.C.: Congressional Budget Office, April 1990.

U.S. Congress. *Farm Commodity Market Performance and Economic Forecasts*, Hearing Before the Joint Economic Committee, Congress of the United States, Washington, D.C.: January 23, 1984.

USDA. *CCC Budget Estimates: An Overview and Preliminary Assessment, Interim Report, Staff Working Group on Review of CCC Estimates*, Washington, D.C.: June 8, 1987.

____. Secretary of Agriculture. *1990 Farm Bill, Proposal of the Administration*, February 1990.

____. Economic Research Service, *The 1990 Farm Act & The 1990 Budget Reconciliation Act: How U.S. Farm Policy Mechanisms Will Work Under the New Legislation*, Washington, D.C.: November 1990.

_____. *Fact Book of Agriculture, 1989*, (Washington, D.C.: 1989).

____. Office of the Inspector General, *National Agricultural Statistics Service, Review of Data Used to Determine Deficiency Payment Rates for Upland Cotton*, Washington, D.C., Audit

Report No. 26600-1-At, September 1990.

____. World Agricultural Outlook Board, ***World Agricultural Supply and Demand Estimates***, Washington, D.C., 1983-1988.

____. ***Scope and Methods of the Statistical Reporting Service***, Misc. Publication No. 1308, September 1983.

____. ***USDA Economic and Statistics Review Panel, Report to Secretary of Agriculture John R. Block***, Washington, D.C.: June 28, 1985.

The WEFA Group, ***Executive Summary: US Agriculture and World Trade***, November 1990.

Winkler, Robert L. and Spyros Makridakis, "The Combination of Forecasts," ***Journal of the Royal Statistical Society***, 146, (1983), pgs. 150-157.

Related GAO Products

Crop Insurance: Inaccurate FCIC Price Forecasts Can Affect Program Costs, (GAO/PEMD-91-28, publication forthcoming).

USDA Commodity Forecasts: Inaccuracies Found May Lead to Underestimates of Budget Outlays, (PEMD-91-24, August 13, 1991).

Short Term Forecasting: Accuracy of USDA's Forecasts and Estimates of Meat Production, Prices, and Inventories, (GAO/PEMD-91-16, May 6, 1991).

Dairy Termination Program: An Estimate of Its Impact and Cost- Effectiveness, (GAO/RCED-89-96, July 6, 1990).

Crop Production: Outlook for Post-Drought Recovery During 1989, (GAO/RCED-89-161BR, June 6, 1989).

AIDS Forecasting: Undercount of Cases and Lack of Key Data Weaken Existing Estimates, (GAO-PEMD-89-13, June 1, 1989).

Immigration: Projected Immigration Under S.448 and Recent Trends in Legal Immigration, (GAO/PEMD-89-12, April 4, 1989).

Electric Power: Issues Concerning Expansion of the Pacific Northwest-Southwest Intertie, (GAO/RCED-88-199, Sept. 14, 1988).

Medical Devices: FDA's Forecast of Problem Reports and FTEs Under HR 4640, (GAO/PEMD-88-30, July 1, 1988).

Welfare Report: Projected Effects of Requiring AFDC for Unemployed Parents Nationwide, (GAO/HRD-88-88BR, May 23, 1988).

USDA's Commodity Program: The Accuracy of Budget Forecasts, (GAO/PEMD-88-8, April 21, 1988).

Immigration: The Future Flow of Legal Immigration to the United States, (GAO/PEMD-88-7, Jan. 8, 1988).

Commodity Futures Trading: Purpose, Use, Impact, and Regulation of Cattle Futures Markets (GAO/RCED-88-30,Nov., 10, 1987).

Retirement Forecasting: Evaluation of Models Shows Need for Information on Forecast Accuracy, (GAO/PEMD-87-6A and 6B, December 31, 1986).

Budget Reductions for FY 1986, (GAO/OCG-86-1, Jan. 21, 1986).

Specific Technical Assumptions Affecting Bureau of Labor Statistics 1985 Employment Projections, (GAO/OCE-85-2, May 20, 1985).

Bureau of Labor Statistics Employment Projection: Detailed Analysis of Selected Occupation and Industries, (GAO/OCE-85-1, April 25, 1985).

The Air Force Can Improve Its Forecasts of Aircraft Spare Parts Requirements, (GAO/NSIAD-85-2, Nov. 19, 1984).

Estimated Employment Effects of Federal Economic Development Programs, (GAO/OCE-84-4, Aug. 15, 1984).

Status of the Statistical Community After Sustaining Budget Reductions, (GAO/IMTEC-84-17, July 18, 1984).

Railroad Employment Projections, (GAO/HRD-83-76, July 15, 1983).

Analysis of Electric Utility Load Forecasting, (GAO/RCED-83-170, June 22, 1983).

Bureau of Economic Analysis Should Lead Efforts to Improve GNP Estimates, (GAO/GGD-83-1, Dec. 27, 1982).

Social Security Actuarial Projection, Staff Study, (GAO/HRD-83-92, Sept. 30, 1982).

Minerals Critical to Developing Future Energy Technologies, Their Availability, and Projected Demand, (GAO/EMD-81-104, June 25, 1981).

Social Security Administration Needs to Continue Comprehensive Long Range Planning, (GAO/HRD-79-118, Sept. 20, 1979).

Use of National Economic Models by Federal Agencies, (GAO/PAD-79- 24, May 16, 1979).

The SRS's Crop Reports Could Be of More Uses to Farmers, (GGD-78- 29, April 13, 1978).

Food and Agriculture Models for Policy Analysis, (GAO/CED-77-87, July 13, 1977).

Long Range Analysis Activities in 7 Federal Agencies, (GAO/PAD-77- 17, Dec. 3, 1977).

What the Department of Agriculture Has Done to Improve Agricultural Commodity Forecasting, (GAO/RED-76-6, Aug. 27, 1975).

Forecast of Postal Service Self-Sufficiency Potential, (GAO/GGD-75- 48, Feb. 20, 1975).

Discussion of Bretschneider/Solenberger Papers

Debra E.Gerald, National Center for Education Statistics

Increasingly, forecast evaluation is recognized as an important part of the forecasting process. These two papers look at different aspects of forecast evaluation. The Bretschneider and Gorr paper seeks to understand the process of forecast evaluation within an organizational context such as the Federal government. The Solenberger paper comments on the procedure used by the General Accounting Office (GAO) to evaluate the forecast accuracy of the Department of Agriculture meat and commodity forecasts.

I will begin with the Bretschneider and Gorr paper which examines Federal forecasting practices. The survey polled Federal forecasters listed in the 1989 and 1990 editions of the Directory of Federal Forecasters in five areas: (1) organizational group and users of forecasts; (2) the forecasting process; (3) political influence; (4) forecast evaluation; and (5) demographic information. The authors acknowledge the limitations of the survey results: its lack of representativeness of Federal forecasters, the low response rate (45%) and lack of confidentiality with respect to responses because of coded survey forms.

While the survey items related to demographic information about Federal forecasters are relatively straight forward, some of the items in the survey related to the forecasting process can be subject to different interpretations by the respondents. For example, the authors state that regression analysis and expert judgement constitute the dominant methodology. Yet among the selection of forecasting techniques listed in the survey, there are three choices with respect to the use of experts and judgement: experts for independent variables; experts for dependent variables; and judgement and quantitative. They ranked in the top half of techniques used by Federal forecasters in their survey. There may be some confusion as to the response to these items depending on how the techniques are applied in the different agencies.

For instance, in a number of Federal agencies, using experts for dependent variables means acquiring survey data from other agencies or private sources. Using experts for independent variables means acquiring forecasts from other Federal agencies or economic forecasting services which are usually based on macroeconomic or structural models.

With respect to the criteria for forecast evaluation, the authors conclude that defensibility of method and reasonableness of assumptions are the most important criteria because they must fit into a political process. But they omit the fact that the attention given to assumptions is a vital aspect of forecasting since unreasonable assumptions can lead to unreasonable forecasts. Also, the authors note that several factors influence criteria in the workplace such as the presence of a formal evaluation process and the level of professionalism and experience. Perhaps, the most interesting observation is that organizational size and its distance, in reporting level, from a higher level such as the Undersecretary's behave in a manner contrary to the authors' expectations. Rather than criteria becoming more important because of its importance to the policy process, the importance of technical criteria prevails. Although the authors dismiss the observation because of survey problems, I contend that this observation is more related to the complexity of the bureaucracy in the Federal Government and the process by which decisions are made by policymakers, a process which is not fully understood by forecasters who provide forecasts to the policymakers. In conclusion, the complexity of Federal bureaucracy and lack of communication between the forecaster and policy maker make it difficult to prepare questions that are understandable across agencies.

The Solenberger paper comments on the seven evaluation questions used by the General Accounting Office (GAO) to evaluate forecast accuracy. However, the paper does not evaluate the methods used and therefore is a starting point rather than an end in itself, since the appropriateness of the technique to the forecasting problem is a vital aspect of forecasting.

This paper addresses the evaluation of agricultural meats and commodity forecasts. The evaluation questions used by GAO are:

1. What is the methodology?
2. Who are the users of the forecast?
3. How accurate are the forecasts?
5. Are the errors reasonable?
6. What are the implications of the forecasts?
7. How can forecasts be improved?

The author gives a warning about making sure variables that are being investigated are themselves important enough to warrant the time needed for a full evaluation. This is important, but such a determination would also require extensive research on the part of GAO to screen the variables.

The author proceeds to comment on the remaining six questions as they relate agricultural forecasts. I will address four of these questions. With respect to users, the author cites that agricultural forecasts are used by private sector decision makers, policy makers in government and to implement government programs involving agriculture. Such a variety of uses and users would underscore the necessity of evaluating the methods.

With respect to measures of forecast accuracy, the author lists the most common measures of forecast accuracy such as percentage error (PE) and root mean square percentage error (RMSPE). Although these measures can typically convey some meaning as to forecast performance, they do not provide information as to the sources of errors. At a minimum, GAO should consider decomposing the mean square errors into its contributory sources of errors to determine if error is due to model misspecification or data variability. However, the author states that finding out why errors occur may not be important. On the contrary, monitoring errors is an important aspect of forecasting. Moreover, it is an activity that may take considerable resources to do adequately.

With respect to the reasonableness of the errors, the author recommends comparing errors to other benchmarks to determine whether smaller errors are possible, such as errors associated with competitive or naive forecasts. However, this is not an end in itself because competitive forecast are more likely to be based on different methodologies and naive forecasts include no factors relative to the forecasting situation.

With respect to improvements in forecasts, GAO needs to go beyond the seven questions. They should identify the sources of errors and include costs associated with forecast improvement. Moreover, they should state whether the increase in accuracy warrants the expense. Of course, such a measure would mean increased costs associated with GAO's evaluation efforts as well.

Comments on Bretschneider/Solenberger Papers (Cont.)

Marshall Kolin, U.S. Postal Service

These are welcome papers for very different reasons. The Bretschneider and Gorr paper raises issues I had not previously considered — and then endeavors to assemble evidence toward their resolution. David Solenberger's paper is about evaluation of the forecasting products of one of the most prolific, venerable and respected set of micro-economic forecasts in the world, those of the USDA.

The Bretschneider and Gorr paper honors this association by endeavoring to draw inferences of Federal Forecasting Practice from responses to a Survey of its membership. The Forecasting Practice important to the paper — we are told — is the relationship of forecast evaluation to categories which appear in a chart labeled Figure 1, data gathered in the survey, and a variety of organizational-forecast practice relationships which the authors' intuition suggest.

I sympathize with the issue and problem which the authors wish to explore.

> "This paper seeks to develop a better understanding of forecasting by beginning with organizational context."

and again, of the Research Question:

> "The principal focus of this paper is to begin to understand the complexity of forecast evaluation in organizations. Many criteria are at work in real organizations and those criteria reflect such diverse factors as mission, organizational culture, resource endowments, etc."

As an econometrician and forecaster who once took anthropology courses with Redfield and Sol Tax and a reading course with Ed Shils in, Theory of Social Organization, I find this approach to the study of forecasting practice appealing. I believe that later drafts of this study will be more accessible to me and I look forward to an opportunity to read them, should the authors so honor me.

Before commenting on the tables which summarize attributes of respondents to the survey instrument (and of their organizations), let me offer a suggestion for improving the information available to this research. The suggestion is that a follow up survey instrument be used to obtain more direct responses to questions of the perceived "mission", "culture", and "econometric sophistication" of the respondent's organization.

My amateur guess of ways to obtain such information in a continuation of the reported survey or its successor follow.

(1), What does the respondent consider to be the "mission" of his forecasting organization? — consider to be the "mission" of the larger organization in which his forecasting organization functions?

(2), What does the respondent think his "chief" (the colleague who runs the forecasting group) views the "group mission" to be? What does the respondent think his chief believes the "mission" of the larger organization to be? (Perhaps respondents' beliefs about the "mission" relate to choice of evaluation method — perhaps coincidence-distance between the respondent's belief and that which is attributed to the chief may turn out to have an interactions with other dimensions which will also be of interest.)

(3), Critical to Forecast Evaluation and the econometric sophistication of the unit will be the educational background and experience (forecasting experience?) of the "chief" and/or that of the individual to whom the "chief" reports. If our authors think their background less important than that of the respondent in its impact upon forecasting practice and the organizational culture, then I disagree with them. Might respondents be asked to report on their perception of the formal training of their "chief" and his boss in econometrics-statistics-forecasting-economics? To report on their "chief's" experience in forecasting administration?

Now, comments about the tables.

(1), The authors' might convey a more coherent picture by revising the tabulation in many of the tables. Table 6 provides a good example. It might be useful and revealing if "0" = NOT APPROPRIATE was not treated as a numeric response in calculating Mean or Median. That is to say, measures of location conditional upon the category being of some relevance (Not Important — Essential) are surely of comparable interest to location measures ranging from "Not Appropriate" to "Essential".

(2) Accuracy criteria in addition to "mean square error" really should be included in the survey instrument. MSE may only be truly appropriate if the loss function on errors of forecast is quadratic. If the loss function is not quadratic, a truly sophisticated evaluation of the errors might well not be of mean square error. Further, Accuracy should probably be measured relative to some set of naive forecasts such as: (a) no change from last period; or (b) last period value plus .5 times the difference between last period and the previous period; or (c), etc.

(3) A clear unambiguous statement of the precise form of the regression equations which yield the statistics presented in Table 11 would be most welcome. Might use of examples to help the reader understand the precise form of the regressions be in order? It appears that a regression is run on a dichotomous dependent variable in Table 9. Why use regression rather than a multiple discriminant function or some more recent technique for maximizing distance in a discriminant space? Is the dependant variable "1" for the 24.5% of respondents who used "formal regular evaluation" or the 36% of respondents who have formal regular plus irregular evaluation process (Table 9)? Were respondents carefully instructed in the distinction between "formal" and "informal" evaluation processes? If so what is the distinction? etc. Is it possible that the dependent variable was formed by scaling the re-

sponses which appear in Table 9 with respective values of "1", "2", through "5"? If so, I would strongly suggest that more meaningful results would be obtained by fitting 5 multiple discriminant functions, one to each dichotomous variable. Imaginative comparison and interpretation of the coefficients in the resultant five functions might be revealing.

(4) Why should anyone care if a regression relationship between a dichotomous dependent variable (or arbitrary scaling of qualitative responses) and some numerically scaled qualitative attributes is "significant at the 10% level"? Is it awkward that these significance levels are the product of calculations based upon the assumption of (more or less) continuously scaled multivariate normally distributed variables and "fixed effect" exogenous variables?

(5) Is the report in table 11 the result of a search procedure?
If so, should we be pleased that more than 10% of the potential relationships are significant at the 10% level? Are more than one in ten significant at the 10% level?

(6) Multiple uses of forecasts — in a truly sophisticated organization — might well imply multiple loss functions and use of different estimation rules to satisfy the multiple uses.

Clearly, I look forward to viewing the next revision of this paper. The issues raised are novel and interesting and the authors are to be sincerely thanked for confronting the question of organization and forecast evaluation relationships with data. I hope these comments will contribute to the insight we can obtain from the confrontation.

Mr. Solenberger's paper is tantalizing. Plausible issues of forecast methodology and evaluation are referenced and very general conclusions are stated. More detailed reference to the data which yielded the conclusions would be welcome. These references might be in the form of examples of gaffs in forecasting practice or suggestions which might guide future forecasting practice.

We are reminded to be particularly attentive to measures of forecast error which are "simple" and easily communicated — "... unless other measures exhibit significantly different results." But we search in vain (almost) for details of the implementation of the evaluation methodology. For example, "what are some of the other measures?" Are there typical cases where we might expect them to yield "significantly different results?"

Forecast errors are categorized as "bias error" or "random error". It would be informative to know why the two (somewhat) different sets of three distinctions suggested by Theil and the Mincer and Zarnowitz volumes are not mentioned in the paper.

How is "bias" measured? Is there a mechanical definition of bias, for example, "arithmetic mean percentage error"? Is there some much more subtle definition which abstracts from error plausibly attributed to such exogenous shocks as unusual weather — viewed as beyond the scope of the forecast? (Is the use of a "trimmed mean percentage error" a reference to trimming on subtle judgmental grounds, or mechanistic dropping of extreme errors?) If subtle definitions are used, what can others who forecast or evaluate forecasts learn from the deliberations and findings of the GAO studies on which Mr. Solenberger reports?

Later in the discussion of forecast errors, "We found many of the errors occur due to reasons which forecasters should have considered." Again, examples and/or generalizations based upon the data which yields this conclusion would be even more valuable than the parsimoniously stated finding.

Also helpful would be more examples of the determination of "reasonableness" and "significance" of forecasting error — in the spirit of the reference to [use of USDA's] " ... policy simulation models to evaluate potential budgetary impacts".

It would be pleasing to know more of the means used to determine whether or not a forecast is worth the effort devoted to its creation. For example, can the "economic value" of harvest forecasts for agricultural commodities be developed from commodity futures market price movements which may follow release of forecast revisions?

The bibliography of "related GAO products" and studies of agricultural forecasts provided with the paper are of particular value. This paper by Mr. Solenberger introduces an area of critical study of forecasting results which must be of interest to practicing forecasters and forecasting organizations. I look forward to a later paper which will summarize with examples and generalizations more closely related to the data studied; i.e., forecasts, the models which produced them, the sources of error, and the means used to assay the relationship between the "value" of particular forecasts and the value of the resources used to produce the forecast.

In summary, these two papers are welcome and stimulating to the imagination in quite different ways. It is a pleasure to have been asked to discuss them.

Using Dynamic Interactions to Aid Forecasts: The Case of Selected Urban/Rural Employment Measures

Ronald A. Babula, Economic Research Service, U.S. Department of Agriculture

Introduction. Although the unemployment rate is a widely-used indicator of economic hardship, it does not completely measure labor market distress. Two groups of workers are not included in the ranks of the unemployed. First, there are discouraged workers, those who have given up looking for a job for they think no job is available, and so are not counted as part of the labor force. Discouraged workers would reenter the labor force and look for a job were economic conditions to improve. Second, there are involuntary part-time workers, who are considered in the labor force and employed, but who would prefer a full-time job. This group is also termed "part-time for economic reasons," since typically their hours are cut back when general economic conditions are unfavorable. Other more commonly studied aggregates include the U.S. civilian unemployed and adjusted unemployment (or adjusted-unemployed workers). Adjusted unemployment — also herein referred to as the nonemployed — includes the unemployed, half of the involuntary part-time workers[1] who would prefer full-time employment, and the discouraged workers.

Of particular interest are the differences between labor market conditions in rural areas versus those in urban areas.[2] Since 1980 the rural unemployment rate has been greater than the urban rate, and rural areas suffer with proportionately more discouraged and involuntary part-time workers. Few studies exist that analyze the labor market behavior of the rural discouraged, involuntary part-time, and adjusted-unemployed workers, and none were located tying employment level movements to movements in these groups. The latter analysis is interesting in that it relates changes in employment levels to changes in labor force movements, the primary of which are movements in and out of the discouraged worker category.

This study will focus on the relationship between employment and (1) discouraged workers, (2) involuntary part-time workers, (3) the aggregate of discouraged workers and one-half of involuntary part-time workers, (4) the U.S. civilian unemployed workforce, and (5) adjusted unemployment (i.e., nonemployed workers). The third category will be referred to here as nonemployed-not-unemployed (NNU), as they are the difference between the chosen definition of adjusted-unemployed (nonemployed) workers less the group officially defined as unemployed.

The goal here is to use data-oriented vector autoregression (VAR) methods to analyze quarterly metro and nonmetro data on the total employed, discouraged, involuntary part-time, NNU, unemployed, and the adjusted- unemployed groups in the U.S., and to discern how these aggregates have historically and dynamically responded to movements in total U.S. employment. Such methods are used to reveal what the regularities embedded in the data have to say concerning how these aggregates have dynamically moved together and interacted historically. More specifically, because the economy now appears to be in a rebound from the recession, I discern how history's long run dynamic trends would "handle" increases in overall employment, in terms of metro and nonmetro responses in the levels of the discouraged, involuntary part-time, NNU, unemployed, and the adjusted-unemployed workers.

In so doing, I address five questions concerning metro/nonmetro responses in these five worker groups from a rise in U.S. employment. First, how long would be required (i.e., what are the sample history's reaction times) for the discouraged, involuntary part-time, NNU, unemployed, and adjusted-unemployed groups to respond to a U.S. employment rise? Second, what would be the directions and the dynamic quarterly patterns of such responses? Third, how long would the sample's dynamic history, on average, have metro and nonmetro responses in the five aggregates from an employment increase endure? Fourth, what would be the strength of response in these aggregates to a rise in U.S. employment? And fifth, what are the similarities and dissimilarities in the metro and nonmetro responses of these groups to a rise in employment? While not forecasts, these dynamic regularities reveal how, on average, groups have historically responded to changes in the U.S. employment levels, and hence provide guides concerning patterns, timing, and size for actual 1990-91 point-forecasts of the five employment aggregates for the present economic recovery. While actual event-specific forecasts of the present apparent recovery may differ in magnitude from the five aggregates' average historical responses to U.S. employment expansion since 1973, such long run average patterns of response are good points of reference for determining which point-forecasts are "reasonable" by dynamic history's standards. And such dynamic aspects of metro/nonmetro worker group response to a workforce expansion are issues frequently not addressed by structural employment models, based on static economic theory.

Answering these questions involves estimating the following five vector autoregressions (VARs), and imposing a rise in U.S. employment on each:

> Model I (discouraged worker model): U.S. civilian total employment (USTOT); metro discouraged workers (METDIS); and nonmetro discouraged workers (NOMETDIS).
>
> Model II (involuntary part-time worker model): USTOT; metro involuntary part-time workers (METPART); and nonmetro involuntary part-time workers (NOMETPAR).
>
> Model III (nonemployed-not-unemployed worker model): USTOT; metro nonemployed-not-unemployed (METNNU); and nonmetro nonemployed-not- unemployed (NOMETNNU).
>
> Model IV (unemployment model): USTOT; metro unemployment (METUN); and nonmetro unemployment (NOMETUN).
>
> Model V (adjusted unemployment model): USTOT; metro adjusted unemployment (METADJ); and nonmetro adjusted unemployment (NOMETADJ).

VAR Econometrics. The above questions concern what occurs to the metro and nonmetro levels of discouraged, involuntary part-time, nonemployed-not-unemployed, unemployed, and adjusted-unemployed workers as a result of a rise in U.S. employment. These issues concern what dynamically happens to these metro and nonmetro aggregates between, and not what happens during, the pre-shock and post-shock equilibria (Bessler). More conven-

tional econometric models, which intensively use economic theory, are better-equipped to handle questions concerning what happens at the static equilibria before and after the shock, here a rise in total employment. Such "structural" econometric models often have little to say about what dynamically occurs between the pre- and post-shock equilibria. Although there are a plethora of studies on employment and also on unemployment, few address the dynamics of employment change and labor force movements. These nonforecast dynamics can provide forecasters of the five modeled employment groups with bounds and parameters useful in guiding the formulation of competing forecasts, and in judging which competing model forecasts of a particular variable are most reasonable.

VAR econometrics is a data-oriented method which imposes a minimal number of a priori theoretical restrictions, so as to permit the dynamic regularities that are in the time-ordered data to reveal themselves (Bessler 1984). Many summaries of the foundations of VAR econometrics, and of the techniques themselves, are in the literature, and need not be presented here (see Sims 1980; Bessler 1984). Consequently, the estimated VAR models are presented directly. Models I through V each take the following form:

$$
\begin{aligned}
Y(t) = {} & a_{0,x} \\
& + a_{x,1}(1)*USTOT(1) + \ldots + a_{x,1}(k)*USTOT(k) \\
& + a_{x,2}(1)*METRO(1) + \ldots + a_{x,2}(k)*METRO(k) \\
& + a_{x,3}(1)*NONMETRO(1) + \ldots + a_{x,3}(k)*NONMETRO(k) \\
& + a_{x,T}*TRD + {}_{x}(t) \qquad (1)
\end{aligned}
$$

x = Models I, II, III, IV, V.

Above, the parenthetical numbers refer to lags 1,...,k; t is the current period t. The Y(t) = USTOT(t), METRO(t), NONMETRO(t), and therefore, there are five VAR models, each comprised of three equations. In models I through V, USTOT represents the U.S. civilian employment level. METRO refers to metropolitan discouraged workers or METDIS in model I; metropolitan involuntary part-time workers or METPART in model II; metropolitan nonemployment-not-unemployment or METNNU in model III; metropolitan unemployment or METUN in model IV; and metropolitan adjusted unemployment or METADJ in model V. The NONMETRO variable refers to nonmetropolitan discouraged workers or NOMETDIS in model I; nonmetropolitan involuntary part- time workers or NOMETPAR in model II; nonmetropolitan nonemployment-not- unemployment or NOMETNNU in model III; nonmetropolitan unemployment or NOMETUN in model IV; and nonmetropolitan adjusted unemployment or NOMETADJ in model V. TRD denotes time trend, and subscripts "x" refer to models I, II, III, IV, and V. Each equation includes a set of three seasonal indicator variables. The regressors with nought subscripts are intercepts, and the epsilon terms are white noise error terms. So there are five VARs, each with three equations of the form in relation 1 above.

Theoretically, k in equation 1 is infinity. Applied work requires that the infinite lag structure be truncated to a lag number that is small enough to be operational and large enough for the residuals to approximate white noise. Yet a universally accepted lag selection procedure does not exist. One choice used with some success is Tiao and Box's (1981) likelihood ratio test procedure. The results (not reported here), conducted at Lutkepohl's (1985) recommended one-percent significance level, suggested a one-lag structure for all VARs except model IV, the unemployment model. A two-lag structure was suggested for model IV.

Quarterly, not-seasonally-adjusted data from the Bureau of the Census, Current Population Survey (CPS) were used. At the time of analysis, data were available from 1973.1 to 1990.4. Note that the period of analysis covers three recessions, 1973-75, 1980, and 1981-82, and their subsequent recoveries. The estimation period ends in 1989.4 to omit influences of the recent recession, for which the dynamic results were generated to characterize. Further, because the CPS incorporated the Office of Management and Budget's 1983 classification of nonmetropolitan counties in the data in 1985.3, I incorporated an indicator variable, valued at zero before 1985.3 and at unity otherwise, in each equation.

In each of the five models, the three equations may have contemporaneously correlated innovations or errors. Failure to correct for contemporaneously correlated current errors will produce impulse responses not representative of historical patterns (Sims 1980). A Choleski decomposition was imposed on each of the five VAR's to orthogonalize the current innovation matrix, such that the variance/covariance matrix is identity. The Choleski decomposition resolves the problem of contemporaneous feedback.

The Choleski decomposition requires a sometimes arbitrary imposition of a Wold causal ordering among the current values of the three dependent variables. My VAR ordering begins with the shock variable and then proceeds on the a priori belief that the sequence represents a causal ordering from USTOT to METRO, and then from METRO to NONMETRO. U.S. employment serves as the first variable in each VAR because the other groups modeled were dwarfed in size by USTOT. Hence USTOT movements should elicit movements in levels of discouraged, involuntary part-time, NNU, unemployed, and adjusted-unemployed workers more than one these five aggregates could be expected to appreciably influence total U.S. employment levels. The metro variable precedes the nonmetro variable in each VAR, since the metro component is far greater than the nonmetro component in size.

Influences of a Rise in U.S. Employment. The impulse response function simulates, over time, the effect of a period-one shock in a variable on itself and on the other modeled variables in the system. Impulse responses are obtained by converting each VAR model into its moving average representation. An moving average representation's parameters are complex, nonlinear combinations of the VAR coefficients.

Figures 1 through 5 present the metro/nonmetro impulse responses in discouraged, involuntary part-time, NNU, unemployed, and adjusted unemployed workers, respectively, from a one-percent rise in employment imposed on models I through V, respectively. The data series were modeled in natural logarithms such that shocks to, and impulse responses in, the logged series are approximate proportional changes in the nonlogged series, and approximate percent changes in the nonlogged series when the impulses are multiplied by 100. A figure's impulse responses are not levels, but rather percent changes, in the nonlogged aggregate levels.

Kloek and Van Dijk's Monte Carlo procedures were applied, and a t-value was obtained for each impulse response. These t-values are used to test the null hypothesis of an impulse being zero, against it being nonzero. Evidence at the 10 percent significance level suggested that the highlighted (solid) impulses in the three graphs are statistically nonzero, and these significant impulses are analyzed in the analyses.

Reaction times and response directions. Metro and nonmetro responses in the levels of all five aggregates responded within three months, that is, during the same quarter, of the USTOT

shock. Metro and nonmetro responses of all five groups declined with increases in U.S. employment, as expected. So given the quarterly history of the modeled series since 1973, the five aggregates can be expected to begin declining within a quarter of the start of the expansion of employment.

Patterns and durations of dynamic response patterns. As national employment rises, the numbers of workers in each of the five groups fall, and these decreases are most pronounced early on in the response period. After several quarters, the strength of the declines in the metro and nonmetro aggregates decay in strength towards zero. The long run dynamics in the sample would have METDIS and NOMETDIS pulsate oppositely with USTOT for 7-8 quarters (see fig. 1); have METPART and NOMETPAR pulsate oppositely with USTOT for 3-4 quarters (see fig. 2); have METNNU and NOMETNNU pulsate oppositely with USTOT for 4-6 quarters (see fig. 3); have METUN and NOMETUN pulsate oppositely with USTOT for 4-5 quarters (see fig. 4); and have METADJ and NOMETADJ pulsate oppositely with USTOT for 4-6 quarters (see fig. 5). The discouraged worker series respond to USTOT for longer periods of time than the other four aggregates, as expected. Involuntary part-time workers are already in the labor force and could more easily move to another category, presumably full-time employment, than can the other groups of nonemployed.

The five figures suggest that in each of the five VARs, the metro and nonmetro impulse response patterns share a number of similarities. The reaction times of each figure's plots are "immediate" or within 3-months of the USTOT shock. Generally the metro and nonmetro impulse patterns take on similar shapes in each of the three figures. In figures 2 through 4, the impulses are most pronounced in absolute magnitude earlier on, and eventually decay towards zero at the longer horizons. This general pattern holds for figure 1, although the discouraged worker impulse pattern accelerates its decline earlier than those of the other figures, before also decaying in strength.

The results of the metro and nonmetro discouraged worker groups are interesting in that they indicate that these groups have historically displayed movement in response to employment changes. The controversy over whether or not discouraged workers as a group should be included in the ranks of the unemployed centers on the issue of whether or not they display a "distinctive attachment" to the labor force [see Stevans (1987)]. Since evidence suggests that there has historically been a noticeable and statistically significant response to an employment change, discouraged workers do appear to be cyclically attached to the labor force. They may be legitimately included in a measure of labor market distress.

Strength of dynamic responses and dynamic multipliers. Dynamic multipliers may be calculated from the impulse response results. Consider figure 1 (generated by model I) as an example. Recall that by a VAR's definition, each of the discouraged worker VAR's equations has a specified number of lags of all three modeled variables: USTOT, METDIS, and NOMETDIS (see equation 1). So a period-one shock (increase) in USTOT places all three quarterly equations into motion. Further, the metro and nonmetro variables are modeled in natural logarithms, such that shocks to, and impulse responses in, these variables are proportional changes in the non-logged series (percent changes in the nonlogged series when the impulses are multiplied by 100). To calculate (for example) the METDIS multiplier from USTOT movements, one sums the 8 statistically nonzero METDIS impulses into a cumulated percent change in the "response" variable; sums the corresponding USTOT impulses into a cumulated percent change in the "shock" variable; and then divides the cumulated response change by the cumulated shock change. This METDIS multiplier resembles an elasticity in being a percent change divided by a percent change; differs from an elasticity in being defined over a multiperiod horizon (8 quarters); and demonstrates the significant percent METDIS change elicited over this horizon per point change in U.S. employment. This multiplier value for METDIS suggests that the sample's dynamics would, on the historical average, have METDIS and USTOT pulsate together for periods of 8 quarters, and that each percent change in USTOT would elicit an 11.9 percent change in METDIS in the opposite direction.

Table 1 suggests that each percent rise in U.S. civilian employment has, on average historically, elicited an 11.9 percent drop in metro discouraged workers and an 9.1 percent drop in nonmetro discouraged workers over 7-8 quarters. Over 3-4 quarters, each percentage point rise in USTOT has been historically associated with a 5.7 percent drop in metro involuntary part-time employment and a 4.3 percent drop in nonmetro involuntary part-time employment. For every point rise in USTOT, the nonemployed-not-unemployed fall 7.3 percent in metro areas and 4.9 percent in the nonmetro areas over a period of 3-4 quarters. And similarly, each point rise in USTOT elicits 7.5 and 6.8 percent declines in metro and nonmetro unemployment, and drops of 7.5 and 6.3 percent in metro and nonmetro adjusted employment.

The five figures and table 1 suggest that, in terms of percent changes, the metro variables respond with slightly more strength than the nonmetro variables. For example, an expansion's increase in U.S. employment elicits more of a percentage decrease in discouraged, involuntary part-time, NNU, unemployed, and adjusted-unemployed workers in metro than in nonmetro areas. The impulse response and multiplier results suggest that the five groups respond with greater strength and for longer periods of time in the metro areas rather than in the nonmetro areas. Each point change generates a greater (absolute) multiplier value for metro series than for nonmetro series. The excess of the metro multiplier (absolute) values over nonmetro values is particularly evident for NNU workers. However, this result may be offset by the fact that nonmetro areas have disproportionately more discouraged and involuntary part-time workers.

Table 1. Dynamic Response Multipliers for a Percent Rise in U.S. Civilian Employment.

	Duration	Nonmetro	Duration	Metro
Discouraged workers:	-11.9	8	-8.1	7
Involuntary part-time workers:	-5.7	4	-4.3	3
Nonemployed-not-unemployed workers:	-7.3	6	-4.9	4
Unemployed workers:	-7.5	5	-6.8	4
Adjusted-unemployed workers:	-7.5	6	-6.3	4

Historical Dynamics in Forecasting Versus Forecasting. Sims (1989) notes that one cannot expect economics to differ from other disciplines and be exempt from having an expanding choice of different model types with varying levels of detail for different purposes. Economic models vary along a spectrum from unrestricted and data-oriented models which only loosely use theory to purely theoretical models with little or no connection to observed events. Sims (1989) and Friedman note that researchers select models along this spectrum based on five

criteria: the degree to which the model incorporates theory; the degree to which the model connects to the data; the confidence levels at which hypotheses are invoked and tested; the analytical purpose at hand; and whether the model predicts acceptably out of sample. Sims (1989) further notes that no model will ever perfectly meet all of these criteria.

The analytical purpose has been to use data-oriented and nontheoretical VAR models which deliberately impose as few theoretical restrictions as possible, so as to permit the dynamic regularities present in the data to reveal themselves. It was then demonstrated how these historical dynamics can be useful to competing forecasters in reconciling particular forecasts, here the ten metro and nonmetro unemployment groups. So rather than focus on the forecasting criterion, I focus on connecting to the data, at the expense of intensity of theory's use, to obtain "complementary" nontheoretical dynamic results of use in guiding the forecasts and/or implied policies of more theoretically based models. For example, if forecaster A's model predicts that in the current apparent rebound, metro discouraged workers will fall by 20 percent, while forecaster B's results suggest that the series will fall by 5 percent, then perhaps the two should start reconciling toward middle ground, since the sample's historical dynamics suggest that the long run average would have discouraged workers fall by 11.9 percent for each rise in U.S. employment. Granted, the present apparent rebound may differ from the average trends in the sample, but insofar as the sample encompasses several past recoveries, then the long run dynamics may provide a good starting point for discussion.

The results also provide complementary information not well-provided by more structural models based on static theory. Forecasters of the 10 series (metro and nonmetro levels of the five modeled worker groups) may note that all series have historically commenced declining within a quarter (i.e., three months) of a U.S. employment expansion. On historical average, a U.S. employment increase has elicited responses which have endured from 3-4 quarters for involuntary part-time workers to as long as 7-8 quarters for discouraged workers. An employment expansion's effects on the five worker groups are felt in the beginning of these response periods, with the impulses being of more pronounced strength early-on in the response cycle, and of decaying strength thereafter. Forecasters may view the "reasonableness" of their point-forecast estimates of a series, say for the current apparent economic expansion, within the context of how far the forecasts' implied change are from the long run average responses implied by the dynamic multipliers. Looking at the relative values of the metro and nonmetro multipliers, in particular that the percent responses are more pronounced in the metro areas, is another use of those results. My purpose has been not to actually forecast metro and nonmetro levels of the five modeled worker groups, but rather to demonstrate how forecasters can use dynamic VAR results along with the forecasts of more structural models to address policy-relevant issues in a more information- intensive manner.

Summary. VAR econometric methods were applied to quarterly data on U.S. employment and various metro and nonmetro groups of jobless and partially jobless to achieve three goals. First, the study reveals how U.S. civilian employment has dynamically interacted with metro and nonmetro levels of these jobless/ partially jobless groups historically. Second, the results demonstrate how these dynamics are useful to researchers who use more theoretically-based econometric models to forecast these very series. And third, I demonstrate that many of the modeled groups of jobless and partially jobless excluded from the official definition of unemployment have a cyclical attachment to the labor force, and could be included with unemployment to form a more comprehensive measure of employment stress.

An increase in employment was found to elicit declines in all ten series. The involuntary part-time groups appeared to have the shortest response duration (3-4 quarters) and discouraged workers, the longest (7-8 quarters). The strength of the response is slightly greater for the metro groups than for nonmetro, and in particular, for metro versus nonmetro involuntary part-time workers.

The dynamic results are also of use to policy makers and point-forecasters who need to predict the five groups of metro and nonmetro workers for such specific periods as the current apparent economic rebound. The multipliers provide a historical average of dynamic response (here declines) to expansions in U.S. employment, and these averages can be used to judge whether individual forecasts are in line, or out of line, with historical response averages. Forecasters may also find this paper's results, which indicate when, how, and for how long the worker groups have, on average, historically responded to U.S. employment expansions, of policy- relevant interest. While not forecasts, the historical dynamics concerning the responses in the metro/nonmetro levels of the five groups are useful in guiding and in complementing actual forecasts of these series.

Finally, the impulse response results and dynamic multipliers suggest that metro and nonmetro levels of discouraged, involuntary part-time, NNU, and adjusted unemployed — groups excluded from the official definition of unemployment — have a systematic and statistically significant attachment to the labor force. Such attachments suggest that they could be included along with unemployment into a more comprehensive definition of joblessness.

Footnotes

[1] One-half of involuntary part-time workers is used for this measure in order to calculate a full-time equivalent measure of employment. The average number of hours worked per week for this group is about 20. ***Employment and Earnings***.

[2] Throughout this paper, the terms "rural" and "urban" are used interchangeably with the terms "nonmetropolitan" and "metropolitan" and the terms "metro" and "nonmetro."

References

Bessler, D.A. "Analysis of Dynamic Economic Relationships: An Application to the U.S. Hog Market." ***Canadian Journal of Agricultural Economics*** 32(1984):109- 24.

Friedman, M. "Methodology of Positive Economics." ***Essays in Positive Economics***, pp. 3-43. Chicago: University of Chicago Press, 1953.

Kloek, T. and H.K. Van Dijk. "Bayesian Estimates of Equation System Parameters: An Application of Monte Carlo." ***Econometrica*** 46(1978):1-20.

Lutkepohl, H. "Comparison of Criteria for Estimating the Order of a Vector Autoregression." ***Journal of Time Series Analysis*** 6(1985):35-52.

Sims, C. "Macroeconomics and Reality." ***Econometrica*** 48(1980):1-48.

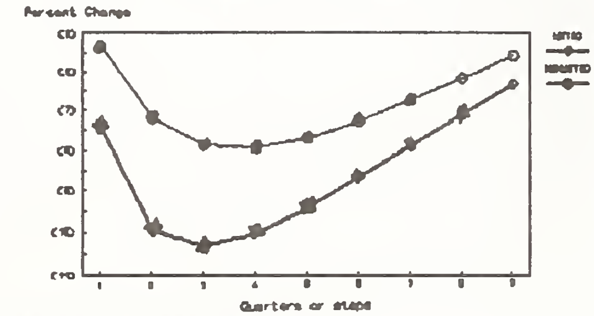
Sims, C. "Models and Their Uses." ***American Journal of Agricultural Economics*** 71(no. 2, 1989):489-94.

Stevans, L., et al. "The Labor Force Attachment of Discouraged Workers." ***Work and Occupations*** 14(February 1987):3-21.

Tiao, G. and G. Box. "Modeling Multiple Time Series with Applications." *Journal of the American Statistical Association* 76(1981):802-16.

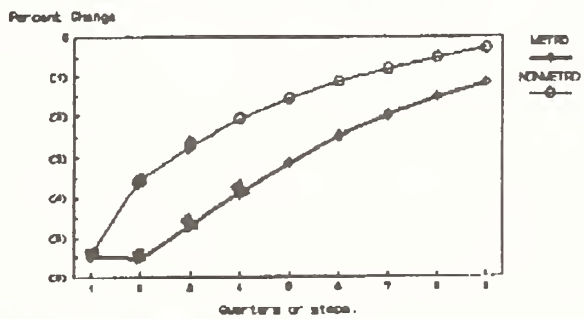
U.S. Department of Labor, Bureau of Labor Statistics. *Employment and Earnings* 38(January 1991).

Figure 1. Model I, Impulse Responses in Discouraged Workers.



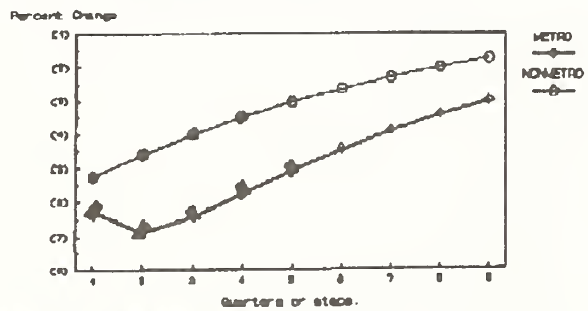
Highlighted (solid) impulse responses are statistically nonzero at the 10% significance level.

Figure 2. Model II, Impulse Responses in Involuntary Part-Time Workers.



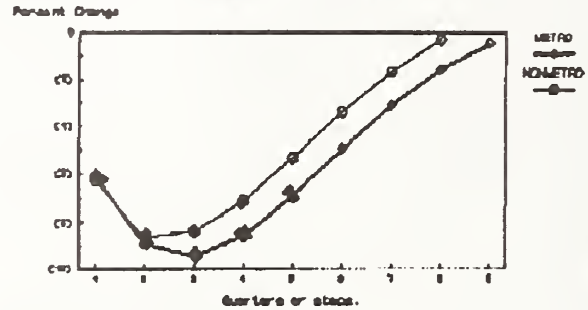
Highlighted (solid) impulse responses are statistically non-zero at the 10% significance level.

Figure 3. Model III, Impulse Responses in Nonemployed-Not-Employed Workers.



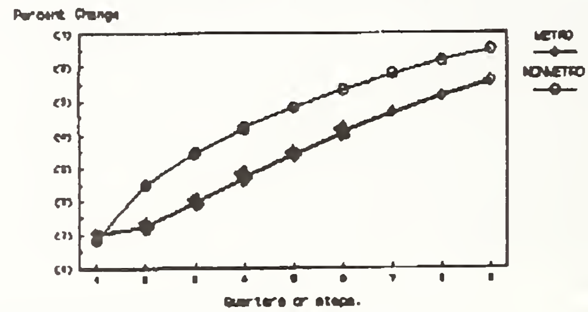
Highlighted (solid) impulse responses are statistically non-zero at the 10% significance level.

Figure 4. Model IV, Impulse Responses in Unemployed Workers.



Highlighted (solid) impulse responses are statistically non-zero at the 10% significance level.

Figure 5. Model V, Impulse Responses in "Adjusted" Unemployed Workers.



Highlighted (solid) impulse responses are statistically non-zero at the 10% significance level.

Using Dynamic Interactions to Aid Forecasts: TheCase of Selected Urban/Rural Employment Measures: A Comment

Frederick L. Joutz, Department of Economics, The George Washington University

Ronald Babula's paper addresses the impact of changes in employment on three categories of distressed workers, the involuntary part-time workers, the discouraged workers, and the unemployed. While he does not use his "model" for direct forecasting purposes, he recommends that the results are helpful in gauging the impacts in labor markets.

The paper has three objectives: First, the dynamic relationships are estimated between employment and the three categories of jobless/partially jobless workers in urban and rural (metro and nonmetro) areas. Second, the model is used to simulate or measure the response to changes in employment on the other variables and estimate the duration of the response. The results can be used as benchmarks for forecasters using theoretically based models. The third objective is to see if an improved measure for unemployment distress can be constructed.

While the econometric model, a VAR, is atheoretical or data-oriented it does have plausible economic arguments. Since the model appears to explain short run responses the migration issue is not addressed. First, the aggregate employment market is considered to be an engine for attracting and releasing participation in the labor force and employment. Second, there may be differences in responses of potential workers depending on whether they live in metro or nonmetro areas. One could argue that metro labor markets offer more employment opportunities and are more homogenous. Thus one might expect a greater response, actually a reduction in involuntary part-time workers and discouraged workers as employment picks up. The empirical results seem to support economic intuition.

My comments will focus on the econometric technique and the omitted forecasting exercise opportunity in the paper.

The data-oriented VAR approach seems appropriate given the stated objectives. Five different 3 variable VARs are estimated and impulse response functions are produced. The different models compare the dynamic relationships between metro and nonmetro categories of discouraged workers, involuntary part-time workers, a composite of the two, unemployed workers, and an aggregate measure of the jobless and underemployed. The data is available quarterly from 1973.1 through 1990.4 in seasonally unadjusted form. The VARs include three seasonal dummies and a time trend.

Given the likelihood for unit roots at zero and seasonal frequencies the trend and even the dummy variables may be inappropriate. (See several articles in Economic Letters. Especially, one by Hahn Shik Lee from the Bureau of Census.) Also, the gurus of the VAR technique, Sims, Litterman, and Doan, seem to argue against the use of trends in most VAR models with economic time series. I will return to the order of integration of series in a moment.

An alternative to the seasonal dummies is to allow for a lag length of the seasonal cycle or four quarters in this case. The author suggests that there is no accepted way of choosing the lag length. My reading of the literature, including the Sims (1980) article, suggests that there is; in fact it is one of the few hypothesis tests one can perform.

The issue involves whether variable x improves the one-step ahead forecast of y. In the VAR framework, this is not a simple single equation F-test, because the effect could be due to another equation in the system. Thus, the hypotheses are cross-equation by construction. The likelihood ratio test is appropriate when the covariance matrix of residuals is not constrained and takes the form

$$(T-c)\,[\log \det\Omega_R - \log \det\Omega_U] \sim X^2_{((R-U)\cdot n)}$$

where T-c is a degrees of freedom correction and the determinants are of the residual covariance matrices in the restricted and unrestricted cases respectively. The restrictions are the number of zero parameters imposed on lags for the n variables.

Also, the RATS package has a routine which decomposes the historical values of the model variables into a base projection and the accumulated effects of current and past shocks. This permits the research to examine whether movements in y were "predictable" from innovations in x a year ago.

A standard exercise when estimating VAR models of this type is to conduct Granger causality tests. If these were done the results were not reported. The tests themselves are good indicators of the dynamic relationships between the variables.

Increasingly the cointegration/error correction model issue arises in macroeconomic studies. The author has indicated he intends to explore them in a future paper. There may be stable long run relationships, such as a ratio of involuntary part-time to full-time workers or discouraged workers to unemployed workers. This is not an exhaustive list, but only a suggestion of where one might direct the research.

Finally, I would like to address the omitted forecasting exercise available in the paper. The sample for estimation purposes ended in the fourth quarter of 1989. The VAR model could have been used to forecast employment figures for 1990 and compared against a univariate ARIMA model and or any of the econometric models like DRI, Wharton, Conference Board, etc.

Structural Models and Some Automated Alternatives for Forecasting Farmland Prices

Karl Gertel and Linda Atkinson, Economic Research Service, U.S. Department of Agriculture

This paper comes from our efforts to improve the forecasting of farmland prices in the U.S. Department of Agriculture. We start with a simple structural model estimated by ordinary least squares. This model gives estimates with desirable statistical characteristics; it has some forecasting capability but also some limitations. Here we report on the first phase of efforts to improve our forecasting capability. In this phase we examine widely available techniques which do not require extensive analyst intervention or expertise. We compare forecasts from the structural model with those from univariate time series, a variable parameter regression model, and results obtained when OLS forecasts are modified by modeling residual errors. We then briefly discuss ideas for the next phase involving more advanced methods.

Farmland Price Trends. Figure 1 shows average price per acre of farmland in the U.S., collected from U.S.D.A. since 1910. The figure shows a slowly rising series with a sharp upturn in the 1970's, a sharp downturn in the 1980's, and a bottoming out and mild upturn in the late 1980's. The most stringent test of forecasting performance we could think of was the capability to forecast these three trend changes since the 1970's.

Figure 2 shows farmland prices transformed into natural logs. Here it is seen that the boom and bust of the past two decades had an antecedent in the decades of the 1920's and 1930's. Unfortunately the data base of our structural model has no memory of this earlier decline since the data series of the explanatory variables begin in 1940.

The Structural Model. As a capital good the price of land is the sum of the expected future net returns to the land, discounted to an equivalent present value. Let P_t be the price per acre of land at the beginning of year t, X_t the net return at the end of year t, and R_t the interest rate at the beginning of year t. Then assuming that returns and interest rates are known :

$$P_t=\sum_{t=1}^{\infty}\frac{X_t}{(1+R_t)^t} \tag{1}$$

If returns and interest rates are constant, the right hand side of equation 1 becomes a geometric progression which sums to the well known capitalization formula :

$$P=\frac{X}{R} \tag{2}$$

The structural equation is derived from equation 2 except that we no longer assume that returns and interest rates are known. Instead it is assumed that farmland market participants form expectations of future values. If X_t^* is returns expected in year t and R_t^* is the expected interest rate then equation 3 becomes:

$$P_t=\frac{X_t^*}{R_t^*} \tag{3}$$

Equation 3 states that price per acre is directly proportional to expected returns and inversely proportional to the expected real interest rate. Price and returns are in real terms at the price level at the beginning of year t. Equation 3 gives a multiplicative relationship which is estimated in logs. Adding a residual term to allow for the stochastic relationship between farmland prices, returns and interest rates:

$$LogP_t=LogX_t^*-LogR_t^*+Logu_t \tag{4}$$

Expected returns and expected real interest rates are modeled as a distributed lag in which expected values are assumed to be a weighted mean of past observed values with weights generally declining from more recent to earlier observations. The particular lag form selected is the rational lag which we adapted from a model employed by Burt [2]. In the interest of brevity, we go directly to the estimating equation of the forecasting model which we explain informally. The derivation of a similar model is given elsewhere [3]:

$$LogP_t=B_1LogP_{t-1}+B_2LogP_{t-2}+B_3LogX_{t-1}+B_4LogR_t+e_t \tag{5}$$

As before, P_t is nominal price per acre at the beginning of year t, P_t and P_{t-1} are farmland prices in the preceding two years, X_{t-1} is returns per acre in the preceding year and R_t is the real interest rate in year t. Since returns for year t are not known at the beginning of the year and were found to be not significant, returns in the preceding year are employed.

The logic of equation 5 and the expected value of the coefficients can be seen by expressing equation 5 in exponential form . Omitting the stochastic error term, equation 5 becomes:

$$P_t=P_{t-1}^{B1}P_{t-2}^{B2}X_{t-1}^{B3}R_t^{B4} \tag{6}$$

According to equation 3, returns and interest rates have equal but opposite effects on farmland prices. Therefore the expected value of $B_4 = -B_3$. Hence equation 6 can be written as:

$$P_t=P_{t-1}^{B1}P_{t-2}^{B2}\left(\frac{X_{t-1}}{R_t}\right)^{B3} \tag{7}$$

In equation 7, price per acre is a weighted geometric mean of past land prices and the capitalized value of past returns. The first and second terms of equation 7 are price per acre, one and two years ago, weighted by B_1 and B_2 respectively. The third term is the capitalized value of last year's returns weighted by B_3. The expected value of the sum of $B_1 + B_2 + B_3 =1$.

Equations 4 - 5 do not contain a constant term since the capitalization model in equation 3 has an implied constant term of 1 which assumes a value of zero in logs.

OLS Results. The statistical profile of fitting equation 5 to ordinary least squares (OLS) is given in Table 1. The equation was fit over three sample periods to see how well it could capture the trend changes discussed earlier. The Durbin h test shows no significant first order serial correlation of the residuals and the asymptotic Breusch-Godfrey test for the 1942-1987 period shows no significant first and second order serial correlation[5]. The values of the coefficients are generally stable over the three sample periods examined and except for returns from 1942 to 1972 are significant at the 95 percent level or higher. The coefficients for returns and interest rates are approximately equal but of opposite sign as expected and the sum of the coefficients relating to returns is close to 1. When fitted with a constant term the value of the intercept is not significant.

The out of sample forecasts were made by using the historical values actually realized for the explanatory variables of returns and real interest rates. This is consistent with agency practice

when preparing long range forecasts for agriculture under different policy scenarios. When forecasts of farmland prices are requested, we are given alternative levels of returns and interest rates. As shown in figure 3, the OLS model did forecast a rise in the upward trend of farmland prices beginning in 1973 but the rate of increase in the trend was underestimated. Similarly the downturn beginning in 1983 is anticipated but greatly underestimated (figure 4). The mild recovery from 1988- 1990 is forecast as a bottoming out of the decline (figure 5). Thus given accurate forecasts of the explanatory variables, the model can predict trend increases and trend reversals, but underestimates them.

Selected Modeling Alternatives. Several alternatives to the OLS model were explored to compare and test their performance. The emphasis was on techniques available in an "automated" form - that is, not requiring a large amount of user intervention or expertise - within existing statistical software packages. Procedures were used from the SAS-[7], AUTOBOX[1] and FORECAST MASTER PLUS[4] programs.

Since the goal was to forecast land prices, we first explored univariate time series models of the land price series; that is, techniques which attempted to predict the future of the series from its past history. The PROC FORECAST procedure in SAS/ETS-provides extrapolation techniques which can quickly and easily generate forecasts in an automatic fashion. Our data were observed annually so we could not model seasonality; that left us with two choices of methods in PROC FORECAST since the remaining Winters methods (multiplicative and additive) are designed primarily for seasonal use: a stepwise autoregressive method that combines a time trend with an autoregressive model and uses a stepwise method to select the lags to use for the autoregressive process, and exponential smoothing, which produces a time trend forecast but allows the parameters used in fitting the trend to change over time (with earlier observations given exponentially declining weights).

In addition, we used an autoregressive integrated moving-average (ARIMA) model that follows the Box-Jenkins technique of time series analysis. ARIMA models also forecast the future of a series based on its past history, using the autocorrelation structure of the data to identify a model. Determining the appropriate structure for such a model and estimating its parameters can be quite complex; the AUTOBOX software package we used accomplishes this task in an automated way.

Table 2 shows some forecasting results from the two extrapolation techniques (with the stepwise autoregressive method labelled "trend") and the ARIMA model and compares them with the observed values for the time periods shown. Note that the univariate models use data beginning in 1910, whereas the OLS results shown use data beginning in 1942. When fit using data through 1972, the univariate models were not able to predict the boom beginning in 1973, and severely underestimated farmland prices for 1982. Similarly, when fit for the period 1910-1982, they missed the decline beginning in 1983 and severely overestimated values for 1987. When fit for the period 1910-1987, the trend and ARIMA models failed to forecast the upturn that occurred. Exponential smoothing did detect the upturn but severely over-estimated the values for 1988-1990. None of the univariate models outperformed OLS but the trend model appeared the least ineffective and was therefore included in graphic comparison and short term forecasts.

Returning to the structural model, it appears theoretically reasonable that one or more of the parameters may be time varying rather than constant, consistent with structural changes in the farmland market. The Forecast Master Plus software package contains a Variable Parameter Regression (VPR) procedure to fit a model whose regression coefficients change over time. Specifically, we used VPR to estimate equations corresponding to the OLS model discussed earlier, where the coefficients for returns and interest rates followed autoregressive processes. Table 3 shows the estimation results for the variable parameter model. Figures 6-8 compare the resulting forecasts with those from OLS, the univariate "trend" extrapolation procedure, and the actual realized values. When fit through 1972, the Variable Parameter Regression performed similarly to OLS but not quite as well. For the other two periods of fit, VPR was somewhat better than the other two methods shown but not by much.

Table 4 evaluates one and two year ahead forecasts from randomly selected years from 1973-1988. For this short term forecasting, the VPR model performed slightly better than OLS, whereas the trend model was a poor third.

We investigated the possibility of modeling the residuals from the OLS fits as per suggestions in Pindyck & Rubinfeld[6]. However, a look at the autocorrelations of the residuals using both the AUTOBOX and a more manual approach in SAS PROC ARIMA indicated that there was not enough structure left over to work with.

We also pursued using transfer function methodology to apply time series analysis techniques to model the land value series as a function of its own past and the values of the independent input series returns and interest rates. Problems encountered in the transition to a new version of the AUTOBOX package caused this process to be less "automatic" than desired and we leave this for future research.

Summary and Conclusions. The question posed was whether the forecasts from a structural model estimated by Ordinary Least Squares can be improved with other techniques from widely available software packages. The answer for our example, a model that is generally consistent with economic logic and with data that are generally consistent with the OLS assumptions, is yes, but not dramatically.

When the structural model was re-estimated to allow for variable parameters there was a gain in forecast accuracy. Yet the variable parameter model is a complement rather than a substitute for the fixed parameter OLS model. The variable parameter model was derived from the same structural model as the OLS model and the forecasts from the variable parameter model are confirmed by the OLS forecasts which predict similar trends. Also the OLS model gives coefficients that are more significant and that are more easily explained to the policy maker.

Both the fixed and variable parameter models generally outperformed a number of univariate models despite the longer sample period available for the latter.

Modeling the OLS residuals resulted in forecast adjustments that ranged from zero to very minor. This is a logical result when there is no strong serial correlation of the residuals.

The next step for improving the econometric forecasts is to look for models that are either more realistic or more efficient in extracting information from the residuals. Although usually applied to a larger data base, the transfer function should be examined. This procedure, in which the structural parameters and the residuals are jointly estimated, could conceivably produce more accurate forecasts than our modeling of OLS residuals. Our colleagues in ERS are exploring the forecasting capabilities of error correcting models which provide parameters of long term equilibrium and a short term dynamic structure [8]. We should

also examine a number of variable parameter models in the hope that at least one of these will capture the boom and bust cycles that periodically occur in the farmland market.

References

1. *AUTOBOX The User's Guide*. Automatic Forecasting Systems Inc., 1986.
2. Burt, Oscar R. ***Econometric Modeling of the Capitalization Formula for Farmland Prices.*** American Journal of Agricultural Economics, Vol. 68, No.1, Feb. 1986.
3. Gertel, Karl, and Patrick Canning. ***Returns, Interest Rates, and Cropland Prices in Selected Regions.*** Tech. Bul. 1778, U.S.D.A., ERS, May 1990.
4. Goodrich, Robert L., and Eric A. Stellwagen, ***ForecastMaster Plus User's Manual.*** Electric Power Research Institute, 1988.
5. Johnston, J. ***Econometric Methods.*** Third edition. New York: McGraw-Hill, 1984.
6. Pindyck, Robert S. and Daniel L. Rubinfeld. ***Econometric Models & Economic Forecasts.*** McGraw-Hill Book Company, 1981.
7. ***SAS/ETS User's Guide, Version 6, First Edition***, SAS Institute Inc., 1988.
8. Tegene, Abebayehu, and Fred Kuchler. ***An Error Correcting Model of Farmland Prices.*** Journal of Production Economics, Forthcoming.

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.

Figure 1

U. S. Average Value Per Acre of Farmland, 1910-90

Value per acre

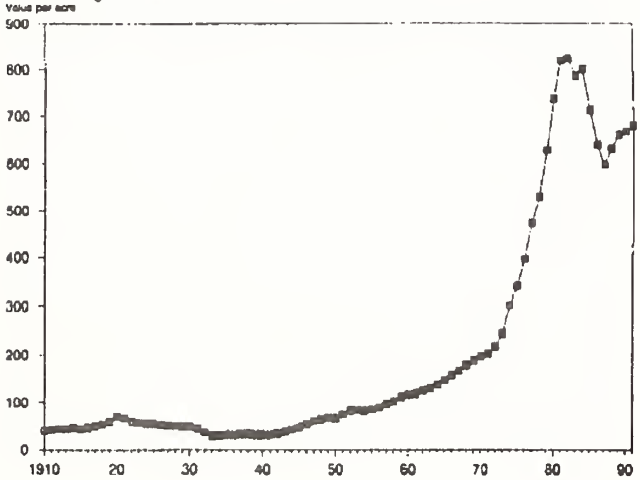


Figure 3

U. S. Average Value Per Acre of Farmland Realized, Fitted, and Forecast by OLS, 1973-90

Value per acre

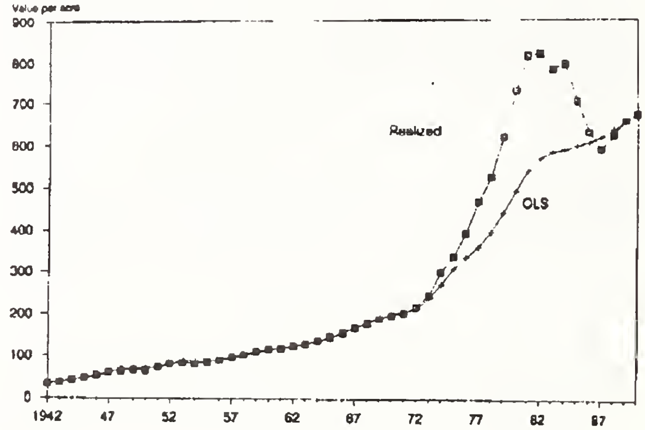


Figure 2

U. S. Average Value Per Acre of Farmland in Natural Logs, 1910-90

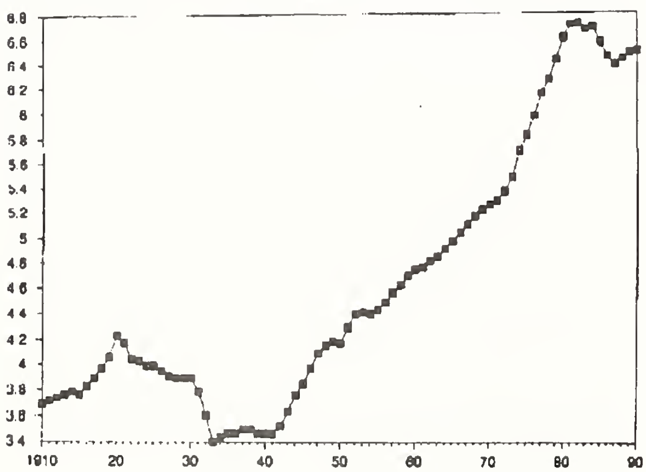


Figure 4

U. S. Average Value Per Acre of Farmland Realized, Fitted, and Forecast by OLS, 1983-90

Value per acre

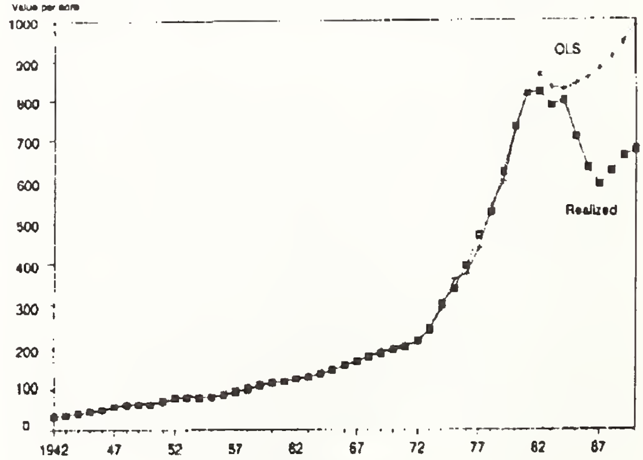


Figure 6

U. S. Average Value Per Acre of Farmland Realized, Fitted, and Forecast by OLS, 1988-90

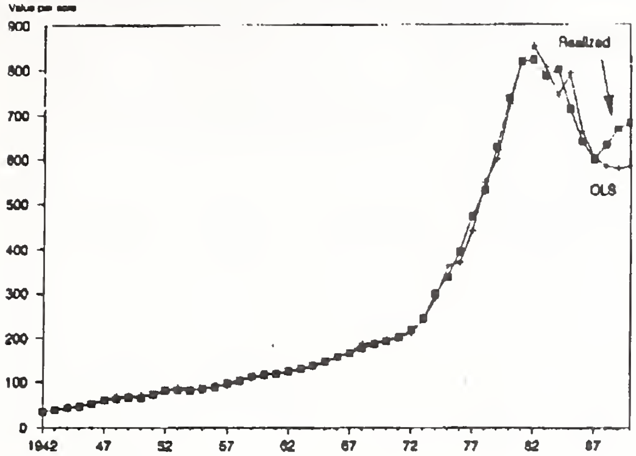


Figure 7

U. S. Average Value Per Acre of Farmland Realized and Forecast by OLS Variable Parameter Regression and Trend with Residual Modeling, 1983-90

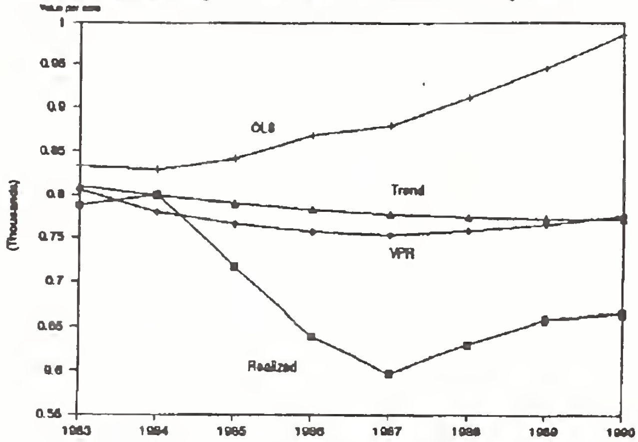


Figure 8

U. S. Average Value Per Acre of Farmland Realized and Forecast by OLS Variable Parameter Regression and Trend with Residual Modeling, 1973-90

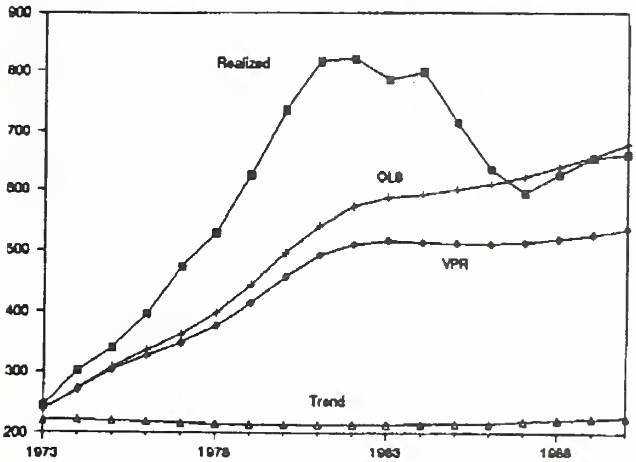


Figure 9

U. S. Average Value Per Acre of Farmland Realized and Forecast by OLS Variable Parameter Regression and Trend with Residual Modelling, 1988-90

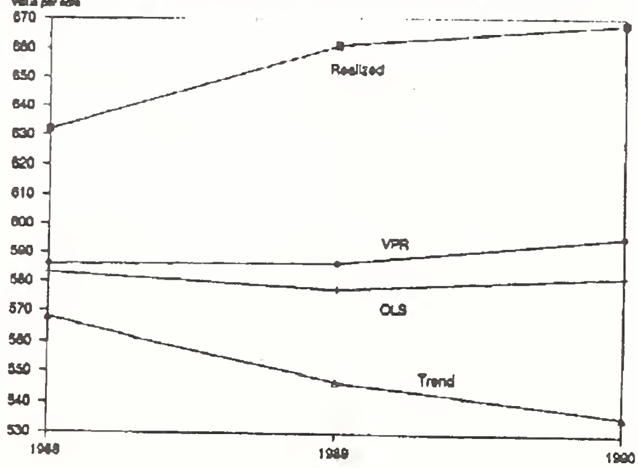


Table 1. U.S. average price per acre of farmland related to lagged farmland prices, returns to assets and the real interest rate, estimated by ordinary least squares.[1,2]

Sample Period / Coefficient	1942-1987	1942-1982	1942-1972
Price of land lagged by one year	1.4279 (12.41)[3]	1.2804 (9.46)	1.2919 (8.64)
Price of land lagged by two years	-0.4676 (-4.24)	-0.3178 (-2.39)	-0.3173 (-2.12)
Returns to assets lagged by one year	0.0389 (2.77)	0.0474 (3.66)	0.0267 (1.43)
Real interest rate	-0.0421 (4.62)	-0.0388 (-4.48)	-0.0317 (-3.10)
Standard error of regression	0.0408	0.0359	0.0320
Sum of coefficient relating to returns	0.9992	1.0100	1.0013

[1]Coterminous U.S.
[2]All variables in natural logs.
[3]Number in brackets are t values.

Table 2 Univariate forecasts of farmland price per acre.

Sample period Forecast year	1910-72 1982[1]	1910-82 1987[2]	1910-87 1988-90[3]
Method	dollars	dollars	dollars
Trend.[4]	214	780	568,547,536
Exponential smoothing[5]	464	1361	1037,1140,1255
ARIMA[6]	278	1555	558,513,472
OLS[7]	544	881	583,578,583
Observed	823	599	632,661,668

[1]The year 1982 was the peak of the boom beginning in 1973.
[2]The year 1987 was the bottom of the decline beginning in 1983.
[3]The years 1988-1990 saw an upturn following the low point in 1987.
[4]SAS STEPAR method of proc forecast. A trend model with modeling of residuals from trend.
[5]SAS EXPO of proc forecast method. Exponential smoothing.
[6]AUTOBOX
[7]For OLS all sample periods begin in 1942.

Table 3. U.S. average price per acre of farmland related to lagged farmland price, returns to assets and the real interest rate, estimated by variable parameter regression.[1,2]

Period of analysis	1942-1987	1942-1982	1942-1972
Coefficient			
Price of land lagged by one year	1.3192 (6.08)[3]	1.2102 (3.87)	1.3135 (4.29)
Price of land lagged by two years	-0.3492 (-1.64)	-0.24446 (-0.78)	-0.3412 (1.11)
Returns to assets lagged by one year[4]	0.0279 (NA)	0.0340 (NA)	0.0262 (NA)
Real interest rate[4]	0.0463 (0.95)	-0.0408 (NA)	-0.0321 (NA)
Standard error of regression	0.0463	0.0430	0.0385
Sum of Coefficients relating to returns	0.9701	0.9658	0.9723

[1]Coterminous U.S.
[2]Forecast Master Variable Parameter Program. All variables in natural logs.
[3]Numbers in brackets are t values.
[4]Value to which cooefficient converges, standard errors unavailable.

Table 4 Short term forecasting performance of alternative models.

Forecast Year	Percent error[1] OLS[3]	VPR[4]	Trend[5]	Mean Square Error[2] OLS[3]	VPR[4]	Trend[5]
One year ahead forecasts						
73	-3.2	-2.9	-11.5	.0010	.0008	.0132
76	-7.9	-10.1	-17.9	.0063	.0102	.0321
77	-8.6	-6.5	-21.1	.0073	.0043	.0447
79	-4.5	-6.8	-20.8	.0020	.0046	.0431
82	7.1	4.5	-3.5	.0050	.0020	.0012
83	5.6	2.2	2.7	.0031	.0005	.0007
86	5.3	4.1	6.3	.0028	.0017	.0039
88	-8.2	-7.6	-10.7	.0067	.0058	.0115
Average absolute error	6.3	5.6	11.8	mean.0043	.0038	.0188
Two year ahead forecasts						
74	-14.2	-14.4	-36.6	.0200	.0206	.1339
77	-21.8	-26.4	-41.5	.0474	.0696	.1721
78	-8.8	-5.8	-37.0	.0078	.0034	.1367
80	-8.5	-15.1	-43.4	.0072	.0228	.1883
83	16.6	10.3	-1.8	.0276	.0106	.0003
84	3.5	-2.6	0.0	.0012	.0007	.0000
87	9.3	8.0	10.9	.0086	.0064	.0118
89	-13.9	-12.3	-19.0	.0192	.0152	.0361
Average Absolute error	12.1	11.9	23.8	mean.0174	.0186	.0849

[1](Predicted-actual)/actual lagged by one year.
[2]Square of 1 above.
[3]Ordinary least squares.
[4]Forecast Master, Variable Parameter Regression.
[5]SAS STEPPAR A trend model with modeling of residuals from trend.

Structural Models and Some Automated Alternatives for Forecasting Farmland Prices: A Comment

Frederick L. Joutz, Department of Economics, The George Washington University

The paper by Karl Gertel and Linda Atkinson, "Structural Models and Some Automated Alternatives for Forecasting Farmland Prices," is an ideal paper for this a forum. The authors appear to be in the initial stages of trying to automate or simplify the tasks required to produce a large number of forecasts for farmland prices. They compare the performance of different types of forecasting models and different software packages against the traditional OLS approach based on economic theory.

My comments will focus two subject areas, the modelling techniques, and the forecast comparisons.

First though the authors are to be commended for presenting the data and discussing the history of the series. It seems that too often researchers attack a data set with high powered econometric and statistical tools without plotting the series first. The ocularmetric approach can often suggest which tools to use, what issues to address and problems to expect, and save headaches later on when the high-powered models don't seem to work. In this case, Gertel and Atkinson might have presented first difference of the data to see if the series might become stationary.

The main issue(s) with respect to the modelling techniques involves structural change and parameter stability. The econometric model follows that of Burt (1986) and specifies the nominal price per acre at the beginning of the year as a function of farmland prices in the previous two years, the returns per acre from the previous year, and the real interest rate. The distribution of the disturbance term should be specified as an exponential in equation 6. All variables are transformed into natural logarithms.

$$P_t = \beta_1 P_{t-1} + \beta_2 P_{t-2} + \beta_3 X_{t-1} + \beta_4 R_t + \epsilon_t$$

The capitalization formula for land values predicts that the expected value of $\beta_3 + \beta_4 = 0$. The particular rational distributed lag and capitalization model imply the expected value of $\beta_1 + \beta_2 + \beta_3 = 1$. Table 1 presents the OLS results for different sub-samples of the model 1942-72, 1942-82, and 1942-87. The two hypotheses are discussed, but not formally tested with the F-test. The authors test for autocorrelation and find none. There is no test for heteroskedasticity, which one might expect given the plots of the series.

Although the authors appear to recognize the potential for structural change, no formal tests were conducted i.e. Chow tests or recursive estimation tests for parameter stability. If the 1970s and 1980s were periods of enormous change in the farmland market, this should be done. In particular, one can see that returns are insignificant prior to 1972, but become significant for the full sample.

The authors use a variable parameter model, VPR, chosen by FORECASTER MASTER. I am not sure how this model is estimated so I cannot comment on it. However, when comparing the model results in Table 3 with those from Table 1, only the price lagged one year appears to be significant. This does suggest specification error in both models.

Two final comments relate to the model specification. First, the lag length of prices is arbitrarily chosen based on the model in Burt (1986). Two lags were specified due to degrees of freedom problems there. The data set used here has 20 more observations. The lag length could be chosen using the well known Akaike and Schwarz criteria. Second, the sum of the two lags suggest that the series may follow a random walk in which case the particular rational lag is not stable.

As mentioned earlier, the OLS econometric model is compared against several models chosen automatically by software packages. The different models include a variable parameter model VPR, ARIMA, transfer function (ARIMA with trend), and exponential smoothing. The univariate models use a data set going back to 1910. While larger sample sizes are preferred to smaller ones, in this case it may not be appropriate, because the time path or "trend" of farmland prices appears to be quite different prior to World War II. Thus, the poor results from these models may be due to the effect of the earlier period on the model(s) chosen and parameter estimates.

The forecast evaluations in the paper are interesting, but poorly organized. It appears the authors are just starting a larger project and still deciding on the evaluation criteria for model and software selection.

For instance, the reasons for choosing the forecast sample evaluation periods and horizons are not stated. They need to decide or explain to the reader whether they are interested in short-term predictions (1-2 years), medium term (5 years). or longer term (10+ years). A common forecast evaluation outcome is that a particular model does not dominate all others at every horizon.

In Table 4 they give the percent error and MSE for three of the models at one and two year horizons during selected years between 1973 and 1989. The comparison of the OLS, VPR, and Trend (actually transfer function) models would be greatly enhanced if they had performed it for the entire period. Familiar tests for bias and dispersion RMSE and AE could be calculated to see who won the "horse race".

One issue they do not consider is the potential gain from combining forecasts. There seems to a revival of interest in these techniques in journals like the ***International Journal of Forecasting***.

Finally, the authors suggest further models might incorporate cointegration and error correction models into the stable of potential candidates. This may prove a promising approach given the potential for farmland prices following a random walk. Also, the error correction models have a nice property in that they can capture both short run and long run dynamics leading to consistent forecasts at different horizons.

In conclusion, the paper by Karl Gertel and Linda Atkinson is an interesting start on a longer term project to evaluate the forecasting performance of different models of farmland prices and the potential for automating the process.

The Use of Dummy Variables and the Computation of Unbiased Predictions, Prediction Errors, and Confidence Intervals in Nonlinear Models

William A. Donnelly, Health Care Financing Administration

Introduction. It is well known that standard regression routines can be used to provide predictions and prediction errors for linear and nonlinear models through use of dummy variables and an augmented dataset—Salkever (1976), Fuller (1980), and Pagan and Nicholls (1984). [1] The Salkever technique is straightforward, as long as the regressand is linear; however, the retransformed predictions are biased if the regressand represents a nonlinear transformation. [2] The papers which develop the statistical theory relating to this prediction technique do not to address the question of transformation bias that is present in nonlinear models. Several authors provide correction approximations for the prediction bias present in the logarithmic model case whenever a retransformation is required of the regressand. These bias-corrections are simply incorporated with the Salkever technique.

The transformation bias in nonlinear models other than logarithmic models is less tractable, since an analytic approximation for the general transformation does not exist; therefore, nonparametric resampling techniques must be used. Fortunately, the bootstrap of Efron (1987, 1981) provides the mechanism for developing the predictions and confidence intervals in nonlinear models. As Efron (1987, p. 173) explains "for more complicated situations like the nonparametric confidence interval [and predictions] problem, Monte Carlo sampling is usually needed to calculate the BC_α [bias-corrected] intervals. How many bootstrap replications are necessary? The answer, on the order of 1,000." What this means is that, if the forecasting model entails a nonlinear transformation of the regressand, then roughly 1,000 bootstrap replications must be run in order to derive an unbiased point-estimate prediction and confidence interval for any particular set of values for the regressors. The bootstrap approach is effectively combined with the Salkever technique to provide unbiased predictions and confidence intervals. And, bootstrapping can be easily accomplished on a microcomputer.

The Salkever variable technique is one way to derive forecast values, and it entails augmenting the observations by "invented" values for the explanatory variables, values for which predictions are being sought. The value for the corresponding dependent variable is set to zero. And, the set of model regressors is expanded to include a dummy variable, one for each of the invented observations. Each dummy variable is constructed to correspond to a single invented observation; this "new" regressor takes on the value of zero for all of the actual observations and a value of minus one for the constructed observation to which the respective dummy pertains. Heuristically, what the technique accomplishes is to maintain the estimate of the original regression plane with the dummy variable compensating for the effect of the additional observation. The amount of the compensation is the predicted value. The replication of the original regression with the augmented dataset plus the inclusion of Salkever variables provide additional parameter estimates in the regression output. The parameter estimate for each of the additional parameters is the prediction for the corresponding invented observation; and the standard error of that parameter is the standard error of that forecast. Bootstrap resampling of a Salkever regression involves replicating the estimation process a sufficient number of times, so as to derive unbiased estimates of the prediction and of its standard error.

Predictions and Prediction Errors [3] The standard multivariate regression model for predictions that is developed is:

$$y = X\beta + u, \qquad y_r = X_r\beta + u_r, \tag{1}$$

where y, X, X_r are $n \times 1$, $n \times k$, and $p \times k$ dimensional fixed matrices of observations, respectively. The X_r's are the invented observations, and, y_r is a $p \times 1$ unobserved matrix. Each y_r has a corresponding zero dependent variable vector, which, in conjunction with the addition minus-one regressors, will provide the predictions for the invented observations. The $k \times 1$ unobserved coefficient matrix, β, represents the preferred model parameters with (u', u'_r) being the unknown random variables satisfying: [4]

$$E\,|\,(u', u'_r) = 0',$$

$$E\left[\begin{pmatrix} u \\ u_r \end{pmatrix}(u', u'_r)\right] = I\sigma^2$$

Rewriting (1) in matrix notation the X's and β's have been extended to include the invented observations and the dummy variables:

$$\begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} X & 0 \\ X_r & -I \end{pmatrix}\begin{pmatrix} \beta \\ y_r \end{pmatrix} + \begin{pmatrix} u \\ u_r \end{pmatrix} \tag{2}$$

The first matrix on the right-hand side of (2) is a fixed matrix. The best linear unbiased estimator of (β', y'_r) is the ordinary least squares (OLS) estimator:

$$\begin{pmatrix} \beta \\ y_r \end{pmatrix} = \left[\begin{pmatrix} X & 0 \\ X_r & -I \end{pmatrix}'\begin{pmatrix} X & 0 \\ X_r & -I \end{pmatrix}\right]^{-1}\begin{pmatrix} X & 0 \\ X_r & -I \end{pmatrix}'\begin{pmatrix} y \\ 0 \end{pmatrix} \tag{3}$$

Equation (3) represents the Salkever (1976) prediction technique as modified by Fuller (1980). Here the original observation matrix, X, has been extended by a set of invented observations, X_r, for which predictions are desired. An additional "zero/minus one" regressor is included for, and corresponding to, each of the invented observations. The estimation of the extended set of regressors on this augmented dataset yields the desired y_r estimates. All of the model regression parameters and statistics, except for the coefficient of multiple determination, remain unchanged. The predictions appear as the respective Salkever parameter estimate, and the standard error of that parameter is the standard of that corresponding y_0 forecast.

This technique is equally applicable to linear and to nonlinear models—Fuller (1980) and Pagan and Nicholls (1984). In the case of linear models unbiased predictions, and the standard error of the forecast associated with those point estimates, are directly available in the regression output, and also for models with nonlinear models if the nonlinear regressand value is the prediction that is sought. If conversely, the natural number prediction of the regressand is the desired value then the technique provides biased estimates, because $[E(Y)]^\lambda \neq E(Y^\lambda)$, where λ is a non-unity value; *i.e.*, a simple retransformation of the prediction of a nonlinear regressand will yield a biased estimate. First, this paper will consider the case of logarithmic transformation bias, when $\lambda = 0$, where a parametric approximation exists. Then the analysis will proceed to a solution to the general nonlinear regressand case, where λ is neither unity or zero. This is where

resort to a numerical, nonparametric procedure is required.

Transformation Bias: Logarithmic Regressand Forecasting models that exhibit a logarithmic transformation bias are of the form:

$$\begin{gathered} \ln Y = f(X, u) \\ \text{where, } \hat{Y}_{\text{biased}} = \exp[(\ln \hat{Y})_{\text{unbiased}}] \end{gathered} \tag{4}$$

this format, Y is the dependent variable, X are the independent variables, and u is the independent and normally distributed error term.

In the equation (4) specification if the errors assume a log-normal distribution then the logarithmic regressand predictions are unbiased. A problem arises when the logarithmic prediction is not the value being sought. That is to say, if the object is determining an estimate of the natural number level of Y then simple exponentiation of the prediction of *ln*Y results in a biased estimate of Y. The simple retransformation of a logarithmic prediction to a natural number value does not suffice. This remains true even though the parameter estimates of the model are unbiased. Since the regressand represents the natural logarithm of the actual observation, any predictions must be reconverted to natural numbers so as to be interpreted. The exponentiated predictions from such a model are not unbiased estimates of the natural-number values, *i.e.*, $E(\ln C) \neq \ln E(C)$. The "naïve retransformation [of the logarithmic model predictions] yields [an estimate of] the conditional median function ... which underestimates the conditional mean"—see Stynes, *et al.*, (1986, p. 95)—of the natural number level of the regressor.

An estimate of the conditional mean, *vis à vis* the conditional median, is the prediction that is desired whenever a summation to the total value is sought. Meulenberg (1965), Goldberger (1968), Mehran (1973), Efron (1981), Duan (1983), and Srivastava and Singh (1989) define and advocate alternative bias-correction approximations for this logarithmic case. These corrections are defined to convert the "expected value of the median" to the "expected value of the mean." Several of these approximations represent "blanket" adjustments that do not differ across the observations space. [5] However, the transformation bias that is present in a nonlinear model does differ across the range of the data, and this necessitates the calculation of a specific correction factor for each prediction. [6] Of these various bias-correction procedures, Efron's "bootstrap" and Duan's "smearing estimate" represent nonparametric techniques. And, while a fixed shift-factor may be defined based upon the standard error of the estimate or at the data centroids—see Meulenberg (1965)—the calculation of minimum variance unbiased estimate (MVUE) adjustment factor for each prediction is straightforward, and is the one to be preferred.

Following the discussion of Goldberger (1968) the MVUE adjustment is:

$$F(w,v,p) = \sum_{j=0}^{\infty} \frac{\left(\frac{v}{2}\right)^{j} \Gamma\left(\frac{v}{2}\right) (pw)^{j}}{\Gamma\left(\left(\frac{v}{2}\right)+j\right) j!} \tag{5}$$

This function is defined by its variance, degrees of freedom, and position in the observation space, and it is evaluated using the property that $\Gamma(n+1) = n\Gamma(n)$. The "pw" term, $\bullet(1 - m)s^2$, is the MVUE log-transformation bias adjustment term, and this is a function of the mean square error estimate of $\sigma_\varepsilon{}^2$ in the regression model and of the mean-value variance, $\sigma^2{}_{mv}$. The definition of m in its general form, the one which differs throughout the observation space, is: $m = X_r(X'X)^{-1}X_r$, where the "r" subscript denotes a particular set of values for the independent variables.

The MVUE of the conditional median function, M, is:

$$F_B e^{b_0 - b_1 x} = M[y \mid x] = e^{\beta_0 - \beta_1 x} \tag{6}$$

Here, F_B is the function evaluated at $w = s^2$, $v = N\text{-}k\text{-}1$, and $p = -\bullet m$. [7] The other variables are: N the number of observations; k the number of regressors; and, when $m = m_{00}$ it is the first diagonal element of the variance-covariance matrix, $(X'X)^{-1}$, from the regression. Reliance upon the single matrix element in the calculation, *vis àvis* the entire matrix defined scalar "m" that is a function of the location of the regressors in the observation space.

Similarly, the MVUE of the conditional mean function, E, is:

$$F_A e^{b_0 - b_1 x} = E[y \mid x] = e^{\beta_0 + \left(\frac{\sigma^2}{2}\right) - \beta_1 x} \tag{7}$$

Herein F_A is the function evaluated at $w = s^2$, $v = N\text{-}k\text{-}1$, and $p = \bullet(1\text{-}m)$. The use of m_{00} represents a blanket adjustment across the data space, and is necessarily a less precise approximation than is one which varies across the observation space. For this reason, one should substitute the estimate of $\sigma^2{}_{mv}$ in the analysis.

The logarithmic transformation bias present in the predictions amounts to the ratio of equation (7) to equation (6). The value, $e^{(\bullet\sigma\bullet)}$, represents a shift in the function. If the objective is to derive predictions couched in levels, values that yield the sum of the observations, then an estimate of the conditional mean rather than the conditional median is required. The bias induced by the logarithmic transformation represents a downward shift in the "true" natural number function of Y. That is to say, a simple retransformation will necessarily represent an underestimate of the conditional mean. A naïve bias-correction that adds $\bullet\sigma\bullet$ to the logarithmic prediction will over-correct for the negative bias that is present because of the convexity of the transformation.

It has been demonstrated by Duan (1983), Stynes, *et al.*, (1986), and Srivastava and Singh (1989) that Meulenberg's (1965) approximation to the MVUE is reasonably precise, and, thus, it should serve in most instances. This means that the analyst need not calculate equation (5), as that is unnecessary, but instead merely use the regressand specific definition for m, $X'_r(X'X)^{-1}X_r$, in the bias-correction formula; namely, $+ \bullet\sigma\bullet - \bullet\sigma\bullet_{mv}$. [8] The degree of the transformation bias in the point estimate, which is a minimum at the data centroids, is a function of the distance of the particular set of explanatory variable values from the data means. The distance of the observation values from the centroid is reflected in the MVUE. In addition, the bias in the naïve retransformation of a nonlinear regressand affects the positioning of the confidence intervals (CI's) about that point estimate. [9]

Predictions are reported as point estimates; *i.e.*, a single value for each set of dependent variable values is given, but those estimates have associated confidence intervals. Inclusion of the confidence interval with the model results provides insight as to the reliability of the prediction. Three standard errors are associated with every regression, and these are: 1) the standard error of the estimate, 2) the standard error of the forecast, and 3) the standard error of the mean-value. In terms of variances, as defined by Christ (1966), the first represents the contribution of the error in the estimation of the intercept term of the relationship. This component of the variance of the estimate is a constant and is, therefore, independent of the values for the regressors in the prediction space. In terms of a logarithmic model this confidence

interval represents a constant percentage above and below the point estimate throughout the observation space. The second component represents the contribution of the error in the estimator of the slope of the regression; and, this is proportional to the distance of the particular regressor's value to its centroid. The third portion is the "contribution of the forecast-period disturbance. The first two terms reflect the random variation of $\upsilon_i, \ldots, \upsilon_n$ [the regressand] in the sample period, and the third term reflects that of υ_{T+n} in the forecast period n (Christ, 1966, pp. 550-551). Christ's notation is:

Rewriting this relationship yields:

$$\hat{\sigma}_F^2 = \hat{\sigma}_\varepsilon^2\left[1+\frac{1}{n}+\frac{(X_r-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] = \hat{\sigma}_\varepsilon^2\left[1+X_r'(X'X)^{-1}X_r\right], \text{ or} \qquad (8)$$

$$\hat{\sigma}_F^2 = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_{mv}^2$$

Reiterating these points, the standard error of the estimate (σ_ε) is a single value over the equation space. The other standard errors, the standard error of the forecast (σ_F) and the standard error of the mean-value, σ_{mv}, differ over the range of the data. These sources of error should be taken into account in the context of a nonlinear model whenever unbiased estimates are being sought for the predictions and for constructing confidence intervals as well. The standard error of the estimate, σ_ε, provides the basis for a naïve retransformation correction, one that will replace the negative bias of the simple transformation with a positive bias. [10] The mean-value and forecast standard errors vary with the values of the regressor observations. Use of the standard error of the forecast provides a bias-correction that is sensitive to the distance of the regressors from their centroids. The calculations of unbiased predictions and their associated confidence intervals for a logarithmic regressand model entail use of the variance-covariance matrix of the regression, $[(X'X)^{-1}]$ and estimates of σ_ε and σ_{mv}. The bias-correction is of the form:

$$\widehat{\ln y_{BC}} = X_0'\beta + \tfrac{1}{2}\hat{\sigma}_\varepsilon^2(1-m), \qquad (9)$$

$$\hat{y}_{BC} = \exp(\widehat{\ln y_{BC}})$$

While the Salkever technique provides a biased estimate of y when the regressand is couched in terms of natural logarithms, the bias-corrected prediction (y_{BC}) may be calculated by adding the variance of the estimate and subtracting one-half of the forecast variance, namely:

$$\widehat{\ln y_{BC}} = \beta_{\ln\hat{y}} + \hat{\sigma}_\varepsilon^2 - \tfrac{1}{2}\hat{\sigma}_F^2 \qquad (10)$$

In a regression, the $\sigma^\bullet_\varepsilon$ is automatically provided and the $\sigma^\bullet_{mv}$ may be calculated from knowledge of the $\sigma^\bullet_\varepsilon$ and the standard error of the forecast value, *i.e.* the Salkever variable standard error. This is accomplished by rewriting the definitional identity for $\sigma^\bullet_F$ as: $\sigma^\bullet_{mv} = \sigma^\bullet_F - \sigma^\bullet_\varepsilon$. Thus, the Meulenberg adjustment log-bias correction is provided by the regression output.

The standard error of the estimate confidence interval, denoted CI_ε, is the one which corresponds to the $\sigma^\bullet_\varepsilon$, and this is invariant over the range of the data. The CI_ε usually is not the most appropriate one to consider when evaluating the prediction accuracy of a regression model. [11] The latter two CI's are hyperbolas that expand in size about the prediction as the value of a particular set of explanatory variables diverges from the centroid of the observation space.

The forecast confidence interval hyperbolas, denoted CI_F, and the mean-value forecast confidence interval, denoted CI_{mv}, are minima at the data centroids, and the distance of a particular set of regressor values from the means of the data is the basis for adjusting the width of these confidence intervals. [12, 13] The farther the regressor values lie from the means of the data, the broader is the corresponding hyperbolic-shaped interval about the predicted value. The σ_F provides the widest confidence interval and the σ_{mv} is the most narrow confidence interval, when measured at the mean of the regressand. The ordering of these confidence intervals, at any given level of statistical significance for a given set of explanatory variable values, is: $CI_F > CI$ and $CI_F > CI_{mv}$. The construction of confidence intervals has been developed elsewhere for the case of models with linear dependent variables—see Kelejian and Oates (1989, pp. 125-128); therefore, that will not be reviewed here. The unbiased confidence interval calculations for a prediction is:

$$CI_\varepsilon = \exp\left(\widehat{\ln y_{BC}} \pm t_\alpha\hat{\sigma}_\varepsilon\right),$$

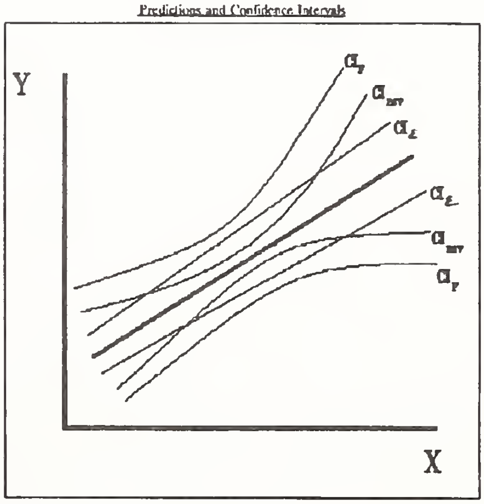
$$CI_{mv} = \exp\left(\widehat{\ln y_{BC}} \pm t_\alpha\hat{\sigma}_{mv}\right), \text{ and} \qquad (11)$$

$$CI_F = \exp\left(\widehat{\ln y_{BC}} \pm t_\alpha\hat{\sigma}_F\right).$$

Figure 1 depicts the three confidence intervals.

Figure 1

Predictions and Confidence Intervals



Example: Logarithmic Regressand A model for Medicare capital-related costs is constructed using data from the Health Care Financing Administration Hospital Cost Report Information System of public-use files. A model of total Medicare capital-related costs per discharge (defined as the total Medicare capital pass-through before reduction)—MEDKDIS—includes four types of bed days available to Medicare patients: 1) routine-care beds (MEDRT), 2) intensive-care beds (MEDIC), 3) coronary-care beds (MEDCC), and 4) other-special-care beds (MEDOSC). These variables are couched in terms of total bed days available per Medicare discharge. Additional explanatory variables are included, so as to control for specific characteristics of the hospital. These include the total beds in the hospital (TOTBEDS), the hospital's occupancy rate (OCCPRATE), the disproportionate-share adjustment (DISPROSH), the hospital's case-mix index (CMI88), and a variable to proxy the AGE of the facilities. This last variable is defined as the total fixed assets divided by the annual depreciation. The model includes three dummy variables, which serve to identify: 1) URBAN hospitals (those located in a Metropolitan Statistical Area), 2) SMall HOSPitals (ones with fewer than 100 beds), and 3) hospitals providing some form of MEDical EDUCation. The descriptive statistics appear in Table 1.

The preferred model with these data is the linear-in-logarithms (log-log) model. The estimated model is thus:

$$
\begin{aligned}
\widehat{\ln y_j} = \ln(\widehat{\text{MEDKDIS}_j}) = {} & \beta_0 + \beta_1 \ln(\text{MEDRT}_j) + \beta_2 \ln(\text{MEDIC}_j) + \\
& \beta_3 \ln(\text{MEDCC}_j) + \beta_4 \ln(\text{MEDOCS}_j) + \\
& \beta_5 \text{URBAN}_j + \beta_6 \text{SMHOSP}_j + \beta_7 \text{MEDEDUC}_j + \\
& \beta_8 \ln(\text{TOTBEDS}_j) + \beta_9 \ln(\text{OCCRATE}_j) + \\
& \beta_{10} \ln(\text{DISPROSH}_j) + \beta_{11} \ln(\text{CMI88}_j) + \\
& \beta_{12} \ln(\text{AGE}_j)
\end{aligned} \tag{12}
$$

Estimation results are presented in Table 2.

In the immediate instance, the mean of the observed dependent variable, MEDKDIS, is \$473.28, whereas the naïve retransformation of the in-sample point estimates from the log-log model yields a mean value of those predictions of \$406.18. Adjustment of the predictions in this log-transformation bias yields a mean of \$468.58 for the retransformed point estimates. The latter number is actually an estimate of the conditional median, that is to say, equation (6). Table 3 presents some of the point estimate *ex post* (in-sample) forecasts that the log-log model specification provides. The 95 percent mean-value and the 95 percent forecast confidence intervals are also reported.

The forecast and confidence interval bounds have been adjusted for the logarithmic transformation bias by using the MVUE procedure as defined in equation (6). For illustrative purposes the initial, middle, and final twenty observations that are reported in Table 3, are from a rank ordering (low to high) of the actual Medicare capital-related costs and are presented reordered by the point-estimate values. It may be noted that in twenty-four (24) of these 61 *ex post* forecasts, the actual cost falls within the 95 percent forecast confidence interval, but in only one (1) of these instances does the actual value fall within the 95 percent mean-value confidence interval. In addition, the preponderance of the hits occur when the actual value is close to the geometric mean of the data (\$375.06).[14] This outcome is not surprising, since a model forecasts best around the data centroids. Models exhibit decreasing precision the farther the regressand from its mean as evidenced by the increasing breadth of the mean-value and forecast confidence intervals. (In twenty-one of the hits registered in Table 3 the actual costs fall within the range of \$380 to \$385.) This suggests that, as the observed values deviate further from the geometric mean of the regressand, the ability of the estimated relationship to provide good point estimates declines.

Table 4 presents some distribution statistics relating to the in-sample forecasts that are generated by the log-log model. The *ex post* forecasts of Medicare inpatient capital-related costs for the 4,265 hospitals are rank ordered by the forecast values, and information on the forecast-value quartiles is given for each of the three confidence intervals. It is seen that in each of the forecast quartiles the mean of the actual cost for that quartile falls within the 95 percent mean-value confidence interval of the forecasts. For example, in the third quartile, the mean of the actual Medicare cost is \$525.3 and this amount is in the CI_{mv} range of \$493.1 to \$551.7, while the average point estimate for that same quartile is \$521.5. The CI_{mv} is consistently the narrowest interval with the CI_F being consistently the broadest interval. For this particular quartile, the CI_F ranges from \$182.5 to \$1,490.2. Also, it should be noted that the standard deviation of the point-estimate forecast in each of the quartiles is always substantially smaller than is the corresponding standard deviation of the actual observed values. This reflects the greater variation that is present in the actual regressand *vis à vis* the preferred model's forecast of the regressand. The forecast values in the first three quartiles exhibit relatively similar variation. The distribution of the forecasts in those quartiles is not markedly skewed. Conversely, the actual regressand distribution is recognizably skewed in each of these forecast defined quartiles. In particular, in the fourth quartile the distribution of the actual observed values exhibits a large positive skew whereas the values of the forecast regressand are negatively skewed.

General Nonlinear Regressand Bias Regression analysis is utilized in mass appraisals to establish consistent estimates of real estate values. The objective of determining the expected value, E(Y), for a property in mass appraisal is adversely affected whenever the dependent variable, the value of the property, is the object of any nonlinear transformation. This results because the $[E(Y)]^2 \neq E(Y^2)$, and while this is a well-know statistical property, it is generally ignored in preparing predicted values. Cassel and Mendelsohn (1985, pp. 137-138) suggest that the nonlinear "transformation introduces a bias which can make it [*i.e.*, the model specification having a nonlinear regressand] inappropriate as a forecasting device for the untransformed variable." Some degree of bias exists in the "naïve" retransformation of the nonlinear estimates, and, those retransformed values are the values that are being sought by the appraiser. Thus, while there is a transformation bias present in the prediction from a logarithmic regressand model, that bias can be eliminated—the parameter estimates themselves are unbiased and do not necessitate correction. A numerical solution to the transformation bias problem is Efron's (1981) bootstrap procedure.

Property appraisals are reported as point estimates, *i.e.* a single value for each property is given, but such estimates have associated confidence intervals. The construction of confidence intervals in real estate analysis has been developed elsewhere for the case of linear regressand models—see Donnelly and Andrews (1988), Epley and Burns (1978), and Janssen (1977); therefore, that will not be reviewed here. The multiple regression model that serves as the basis for mass appraisals is founded upon the concept of an hedonic price, a model that postulates the value of a property, Y_j, as being a function of the physical characteristics of the property, X_{ij}, the environmental attributes associated with it, the financial exigencies of the market in general, and of the buyer and of the seller specifically, Z_{ij}—see Edmunds (1984). In this ex

ample, which is presented for expository purposes, only the physical attributes of the property which affect the value, the X_{ij}'s, are considered in a general mathematical specification of the relationship, namely:

$$Y_j = f(X_{ij}) + \epsilon_j \tag{13}$$

where Y_j = value of the j^{th} property
X_{ij} = i^{th} physical characteristics
$\varepsilon + N(0,\sigma^2)$ & $cov(\varepsilon_i,\varepsilon_j) = 0$, for $i<>j$, error term

There is no *apriori* reason to expect that the relationship defined in equation (13) should assume any particular mathematical form. The question as to what the exact form of the relationship is in the case of hedonic models has engendered a lively debate in the literature. Recently, attention in the appraisal literature has been directed toward evaluating the most appropriate functional form needed to generate the estimates of the property values—see Murphy (1989). Generalized testing the functional form of the relationship is accomplished through the application of the Box-Cox transformation procedure. A brief exposition on the Box-Cox transformation will suffice here.

The Box-Cox Transformation. A general model specification, one that relies upon a flexible functional form based on the Box-Cox (1964) transformation procedure, may be applied to equation (13). This is a means of addressing the question of whether specific attributes contribute to the determination of a property's value in a linear, or in a nonlinear, fashion. The Box-Cox procedure is:

$$Y_j^{(\lambda_0)} = \frac{Y_j^{\lambda_0} - 1}{\lambda_0}, \text{ when } \lambda_0 \neq 0 \qquad = ln(Y_j), \text{ when } \lambda_0 = 0 \tag{14}$$

The superscript notation "(λ_0)" on the left-hand side of equation (14) identifies a Box-Cox transformation. This involves converting the actual value of the variable by either raising that variable to the λ_0 power or by taking the (*ln*), natural logarithm. (The nature of the transformation requires the Y_j's to be strictly positive.) The specific nonlinear transformation depends upon the estimated value for the λ_0 parameter. Values for λ of either 'unity' or 'zero' provide the most familiar functional forms, but other forms are also derivable. The means of evaluating alternative values of the λ's is to consider the logarithm of the likelihood function of the regression.[15] If the estimate of λ is equal to unity the simple linear model is suggested by the data.[16] A value for λ equal to zero implies that the natural logarithms of the data should be taken.

The generalized linear Box-Cox (GLBC) functional form represents a specification of equation (13) encompasses many other functions. The GLBC is:

$$Y_j^{(\lambda_0)} = \beta_0 + \sum_{i=1}^{n} \beta_i X_{ij}^{(\lambda_i)} + \epsilon_j \tag{15}$$

It should be noted that provision in this specification is made for highly nonlinear models. The SHAZAM® package provides a convenient implementation of the Box-Cox estimation procedure that provides maximum likelihood (MLE) estimates of the β's and λ's—see White, *et al.* [1990]. The functional form is evaluated by comparing the value of the logarithm of the likelihood function (LLF) for the unrestricted model to the LLF of the restricted model—see Savin and White (1978). The unrestricted model is the GLBC against which the performance of the more restricted models are gauged. This is accomplished by making pair-wise comparisons between those models and the GLBC; for example, the number of λ's that must be restricted to unity in order that the GLBC exhibit the linear model is one for the dependent variable and one for each of the regressors in the model. (Dummy variables must be excluded from the transformation because by construction these variables are not strictly positive.)

Example: GLBC Model The data on selling prices and characteristics of properties used in the analysis are from a local tax assessor's office. The database represents a sample of 325 observations that include houses selling in the price (SELPRC) range between \$13,100 and \$280,000 with the median price of \$48,000. The median floor space (TOTFLR) is 1,324 square feet, and the median value for the construction index (CONSI) is 1.0. The properties selected for analysis represent more than 2• percent of those on the local tax rolls. The STHSIDE regressand is a dummy variable that takes on the value of 'e'—the base of the natural logarithms, *i.e.* e = 2.71828..., if the property is located in a neighborhood on the south-side of the city and a 'one' if the property is located on the north-side of the city.[17] Information on a total of fifteen (15) variables was collected on each of the properties, but only six (6) variables are used in this analysis. Table 5 presents the summary statistics relating to the model data.

The data are standardized by dividing each variable by its respective mean. The explanatory variables are: the square footage of living space, the number of BATHroomS, the AGE of property, the index of quality of the original construction—and the TOTal LaST assessed value before the most recent sale. The recent selling price is the regressand in the model. The dummy variable representing the location of the property is also included in the specification. The functional form analysis suggests that the generalized Box-Cox (GLBC) and the linear Box-Cox (LBC) models provide results that are statistically different from the standard alternative specifications tested, such as the linear or linear-in-logarithm forms. Estimation results for the three Box-Cox transformed dependent variable specifications are presented in Table 6. Of these forms the LBC model presents the best F-statistic value, the next best R^2 value, the smallest standard error of the estimate; and it is the most parsimonious of these models. Based upon the likelihood ratio test The LBC form is not statistically different from the GLBC. Therefore, the LBC is adopted as the preferred model.

Specifically, the preferred LBC real estate model is:

$$\hat{Y}_j^{(\lambda_0)} = \left(\frac{\left(\frac{SELPRC_j}{\mu_{SELPRC}}\right)^{\hat{\lambda}_0} - 1}{\lambda_0} \right) = \beta_0 + \beta_1 \frac{TOTFLR_j}{\mu_{TOTFLR}} + \beta_2 \frac{BATHS_j}{\mu_{BATHS}} + \beta_3 \frac{AGE_j}{\mu_{AGE}} + \beta_4 \frac{TOTLST_j}{\mu_{TOTLST}} + \beta_5 \frac{STHSIDE_j}{\mu_{STHSIDE}} \tag{16}$$

The predictions of Y_j that the preferred LBC functional form provides are not unbiased estimates because of the nonlinearity of the dependent variable, *i.e.* $\lambda_0 = 0.57$. That is to say, a naïve retransformation of the predictions from this LBC model does not provide unbiased expected values of the selling price. Thus for appraisal purposes, since the regressand in the preferred model is subject to a nonlinear transformation, any predictions prepared from this model necessarily must be converted back to natural numbers. Correction for the transformation bias that is present in the predictions and for the bias in the corresponding confidence intervals must be made. Unfortunately, the transformation bias correction factor for the LBC (and for the other nonlinear but non-logarithmic model forms of the GLBC specification) is not an analytically tractable one; and, therefore, one must resort to a numerical method so to be able to derive the requisite correction factor. This may be accomplished via the procedure known as "bootstrap."

Predictions and Standard Errors: General Nonlinear Regressand In the case of the general nonlinear regressand model specification, Monte Carlo procedures are required to construct unbiased predictions. Salkever's technique when combined with the bootstrap (Efron, 1981) is a method for accomplishing this and for deriving unbiased nonparametric standard errors and confidence intervals, as well. Resort to a nonparametric procedure, such as the bootstrap, is necessary in instances where analytical solutions are either highly complicated or for which those solutions do not exist. The degree of transformation bias that is present and the parameters of the cumulative density function are unknowns for the general nonlinear transformation of the regressand. A practical solution to the problem is the bootstrap. [18] Efron's explanation of the derivation of an unbiased estimate of _, assumes:

"the true standard error of θ is a function of F ... [and] knowing n and the form of θ, the true standard error is only a function of the unknown distribution F.

$$\sigma(F, n, \hat{\theta}(\cdot,\cdot,\ldots,\cdot)) = \sigma(F) \qquad (17)$$

The bootstrap estimate of the standard error, σ_B, is simply

$$\hat{\sigma}_B = \sigma(\hat{F}) \qquad (18)$$

where F is the empirical probability distribution

$$\hat{F}\text{: mass } \frac{1}{n} \text{ on } x_i, \quad i = 1, 2, \ldots, n.\text{" (Efron, 1981, p. 140.)} \qquad (19)$$

But, the function $\sigma(F)$ is not a tractable one and, therefore, Monte Carlo simulations must be used—these are the bootstrap replications. A bootstrap replication entails repeated trials that are drawn with replacement from which summary statistics are calculated. Between 50 and 200 bootstrap replications should provide the unbiased standard errors. The replication process is to:

"*Step 1*. Construct F as at [(19)].

Step 2. Draw a *bootstrap sample* from F,

$$X_1^*, X_2^*, \ldots, X_n^* \overset{iid}{\sim} \hat{F}, \qquad (20)$$

and calculate $\theta^* = (X_1^*, X_2^*, \ldots, X_n^*)$.

Step 3. Independently do Step 2 some number B times, obtaining *bootstrap replications* $\theta^*(1), {}^*(2), \ldots, \theta^*(B)$, and calculate

$$\hat{\sigma}_B = \left[\sum_{b=1}^{B} \frac{[\theta^*(b) - \theta^*(\cdot)]^2}{B - 1}\right]^{\frac{1}{2}}, \qquad (21)$$

where $\theta^*(\cdot) = \Sigma\theta^*(b)/B$.

As $B \rightarrow \infty$, the right-hand side of [(21)] converges to (F). In practice, the author has found B in the range 50-200 adequate for estimating standard errors." (Efron, 1981, p. 140)

A similar procedure pertains to deriving unbiased predictions, and the bootstrap predictions and prediction standard errors can be obtained from application of the Salkever technique. That is to say, the augmented dataset and the extended set of regressors appraisal model requires the bootstrap iteration 1,000 times so as to obtain unbiased predictions and confidence interval estimates (Efron, 1987, p. 173). Deriving unbiased confidence intervals is thus orders of magnitude more computationally intensive than for constructing unbiased standard errors. [19] Thus the modeling strategy is to bootstrap the Salkever technique and use the Salkever variable predictions and standard errors to derive unbiased estimates of the retransformed regressand. This bootstrap with Salkever variables may be accomplished in SHAZAM®. [20] (Srivastava and Singh, 1989, provide an example of bootstrapping multiplicative models so as to obtain a confidence interval for the constant term of a Cobb-Douglas model. Naturally, the more general procedure described here will accomplish the same ends; and, in addition, it provides unbiased predictions.)

In the immediate instance, the mean of the observed dependent variable, SELPRC, is \$51,762; whereas, over the observation space, the simple retransformation of the *ex post* forecasts from the LBC model yields a mean value of \$51,535 versus the \$51,767 figure given by the bootstrap procedure. Thus, the bootstrap average better approximates the expected value of the mean. The bias in the point estimates is on average roughly only about -• percent. In Table 7 the bias-corrected predictions and 95% forecast confidence intervals are reported for the first, middle, and last twenty data points. It can be seen that this nonlinear forecasting model appears to be quite robust in its ability to successfully capture the actual observed selling price of a property within the 95% forecast confidence interval. Naturally, it performs relatively better close to the centroid of the observation space in regard to the middle twenty observations. In the tails of the data between fifteen and twenty percent of the selling prices fall outside the defined forecast confidence interval. These confidence intervals should not be taken to reflect unequivocal limits. For example, the property in Table 7 that sold for \$19,000 falls exactly on the 95% CI_F lower-bound and thus is considered to be captured successfully by the model, whereas, the property that sold for \$22,000 falls just below the 95% CI_F lower-bound and is considered to be missed by the model. Any such confidence intervals should be taken to be indicative and not absolute. It is interesting to note that the percentage naïve transformation bias that is measured is larger for the confidence interval lower bound than for either the point estimate or the CI upper bound.

Table 8 presents the bootstrap predictions and 95% forecast confidence interval summary statistics by quartile. These data are rank ordered by the predictions. It is seen that the standard deviation in the predictions is smallest for the second and third quartiles; $1,973 and $3,839, respectively, and is greatest in the tails of the distribution—it is $34,015 for the fourth quartile. The bootstrap mean-value predictions overall and for each of the quartiles are similar to the observation mean-values for those ranges.

Summary This paper has considered the problem of the transformation bias that is present in nonlinear forecasting models. The problem of transformation bias in nonlinear models is well known to forecasters, as is the Salkever dummy variable approach for constructing predictions and standard errors of the forecast. The bias correction within the Salkever formulation has not been previously demonstrated for either the logarithmic regressand or in the more general nonlinear regressand cases. An application of the analytic bias correction necessary for a logarithmic regressand case is demonstrated, which requires a simple calculation using the information provided on the standard error of the estimate and the standard error of the forecast. The numerical generalization to bias correction that is necessary for other nonlinear, but non-logarithmic, regressand models is illustrated using the Efron's bootstrap resampling technique. The bootstrap is computationally intensive, yet tractable on a microcomputer. The resulting bias-corrected predictions and confidence intervals are the ones that forecasters seek; and, therefore, it is recommended that analysts adopt the simple procedures discussed above.

References

Box, G.E.P. and D.R. Cox, 1964, An Analysis of Transformations, *Journal of the Royal Statistical Society* B-26, 211-243.

Cassel, E., and Mendelsohn, R., 1985, The Choice of Functional Forms for Hedonic Price Equations: Comment, *Journal of Urban Economics* 18, 135-142.

Christ, C.F., 1966, *Econometric Models and Methods*, (John Wiley & Sons, New York).

Donnelly, W.A., and Andrews, R.L., 1988, Understanding and Using Variance in Regression Based Appraisals, *The Real Estate Appraiser and Analyst* (Fall), 50-58.

Duan, N., 1983, Smearing Estimate: A Nonparametric Retransformation Method, *Journal of the American Statistical Association* 78, 605-610.

Edmunds, R.G., 1984, A Theoretical Basis for Hedonic Regression: A Research Primer, *Journal of the American Real Estate & Urban Economics Association* 12, 72-85.

Efron, B., 1987, Better Bootstrap Confidence Intervals, *Journal of the American Statistical Association* 82, 171-185.

—, 1982, *The Jackknife, the Bootstrap and Other Resampling Plans*, Conference Board of Mathematical Sciences and the National Science Foundation, (Society for Industrial and Applied Mathematics, Philadelphia).

—, 1981, Nonparametric Standard Errors and Confidence Intervals, *The Canadian Journal of Statistics* 9, 139-172.

Epley, D.R., and W. Burns, 1978, The Correct Use of Confidence Intervals and Regression Analysis in Determining the Value of Residential Homcs, *American Real Estate and Urban Economics Journal* 6, 70-85.

Fuller, W.A., 1980, The Use of Indicator Variables in Computing Predictions, *Journal of Econometrics* 12, 231-243.

Goldberger, A.S., 1968, The Interpretation and Estimation of Cobb-Douglas Functions, *Econometrica* 35, 464-472.

Janssen, C.T.L., 1977, Confidence and Prediction Intervals in Regression Appraisal, *The Real Estate Appraiser* September/October, 25-33.

Kelejian, H.H., and W.E. Oates, 1989, *Introduction to Econometrics: Principles and Applications*, Third Edition, (Harper & Row, New York).

Mehran, F., 1973, Variance of the MVUE for the Lognormal Mean, *Journal of the American Statistical Association* 68, 726-727.

Meulenberg, M.T.G., 1965, On the Estimation of an Exponential Function, *Econometrica* 33, 863-868.

Murphy, L.T., 1989, Determining the Appropriate Equation in Multiple Regression Analysis, *The Appraisal Journal* October, 498-517.

Pagan, A.R., and D.F. Nicholls, 1984, Estimating Predictions, Prediction Errors and Their Stand Deviations using Constructed Variables, *Journal of Econometrics* 24, 293-310.

Salkever, D.S., 1976, The Use of Dummy Variables to Compute Predictions, Prediction Errors and Confidence Intervals, *Journal of Econometrics* 4, 393-397.

SAS Institute, Inc., 1988, *SAS/ETS® User's Guide, Version 6, First Edition*, (SAS Institute, Cary, NC).

Srivastava, M.S., and B. Singh, 1989, Bootstrapping in Multiplicative Models, *Journal of Econometrics* 42, 287-297.

Stynes, D.J., G.L. Peterson, and D.H. Rosenthal, 1986, Log Transformation Bias in Estimating Travel Cost Models, *Land Economics* 62, 94-103.

White, K.J., S.D. Wong, D. Whistler, and S.A. Haun, 1990, *SHAZAM®: User's Reference Manual Version 6.2*, Second Edition, (McGraw-Hill, New York).

Endnotes

[1] Throughout this paper, the technique for constructing predictions and prediction errors will be referred to as the Salkever technique. The approach is variously referred to as the dummy, indicator, or constructed variable approach to computing predictions and prediction errors.

[2] Statistical packages, such as SAS Version 6.0, provide unbiased predictions and confidence intervals for linear regressand, models but these packages do not attempt to correct (analytically or numerically) for transformation biasing.

[3] This section is based upon the work of Salkever (1976), Fuller (1980), and Pagan and Nicholls (1984).

[4] Fuller (1980) generalizes the characterization of the error term with a partitioned Ω matrix so as to provide for heteroscedastic and/or serially correlated cases. For simplicity of exposition, we assume an homoscedastic and serially independent error.

[5] For example, the naïve log-bias correction ($+\bullet s^2$), Meulenberg's MVUE approximation [$+\bullet (1-m_{00}) s^2$], and Duan's smearing correction ($+1/N \Sigma_i e^i$) are all constants across the observation space and independent of the particular values for the regressands.

[6] As will be demonstrated the log-transformation bias-correction factor is, in its derivation, a function of the standard error of the forecast, the standard error definition that differs across the observation space. The standard error of the forecast value is necessary in deriving the forecast confidence interval about the prediction value.

[7] Meulenberg (1965) assumes $m = m_{00} = a'(X'X)^{-1}a$, for $a' = (1, 0, ..., 0)$.
[8] This is the definition for $\bullet(1-m)\sigma^2$ that Srivastava and Singh (1989, p. 292) recommend and that Stynes, *et al.* (1986, p. 99) find, quite incorrectly, "greatly complicates ... analysis."
[9] A confidence interval represents the probability that at some predetermined level of statistical significance the actual observed value will occur within the range of values so defined. That is to say, at the 95 percent level of confidence, the true value is expected to occur within, for example, the range of $430.10 to $476.90. Naturally, the OLS point estimate itself always falls within all of the confidence intervals that may be defined, but this is not necessarily true for the actual observed value for the dependent variable. Thus, the confidence interval, be it narrow or broad, indicates how certain the analyst is that the true value falls nearby the forecast value. The more narrow the CI the more precise is the OLS model in its forecasting abilities and the more meaningful the point estimates provided by it.
[10] This is the "exp $(\sigma^2/2)$" value that the SAS/ETS® *User's Guide* discusses (1988, pp. 78-79).
[11] The reason that the CI_ε is the one CI usually included along with the forecast value on the infrequent occasion when a CI is reported, is that most regression programs present the standard error of the estimate statistic. SAS® will provide the CI_F and CI_{mv} values.
[12] Recall that a regression line (or a regression plane) passes through the means of the data. That is to say, one point on any regression line (or regression plane) represents the coordinates of the mean values of the data.
[13] By "distance" is meant how far, numerically, a particular observation is from the mean value for the variable; *e.g.*, if the mean of Medicare capital-related costs for a sample of hospitals is $473.28, then a hospital with capital-related costs of $777.00 is a greater distance from the mean than is a hospital with capital-related costs of $381.00.
[14] Recall that the mean of logarithmic transformed data is the geometric mean of the original data.
[15] The name "likelihood function" is given to the joint probability distribution of a sample. A class of estimators, that are referred to as maximum likelihood estimators (MLE's), have as their objective function the selection of the parameters in the model that will maximize the joint probability of the sample; that is to say, the MLE provides those parameters that would generate the observed sample most often. A comparison of the natural logarithm of the likelihood function value for different equation specifications can be used in evaluating the superiority of the alternative functional forms. This is particularly useful when the functional forms are nested within a specification such as is the case in the GLBC. The likelihood function that is being evaluated for the GLBC model is:

$$LLF(\beta,\sigma^2,\lambda_y,\lambda) = -\bullet N\, ln 2\Pi - \bullet N\, ln\sigma^2 - 1/\{2\sigma^2[y^{(\lambda 0)} - x^{(\lambda i)}\beta]'[y^{(\lambda 0)} - x^{(\lambda i)}\beta]\} + [\lambda_0 - 1]\Sigma_j\, lny$$

[16] For a $\lambda_0 \equiv 1$, then the Box-Cox transformation is merely $V_j^{(\lambda 0)} = V_j - 1$ which, since it represents a simple linear transformation of V_j, only affects the MRA constant term and does not affect the MRA slope term.
[17] Normally, dummy variables take on the value of one or zero; however, in nonlinear estimation it is often convenient to use "e" and "one" as the values instead. The logarithmic transformation of the unity and zero dummy variable values obtains the intended zero and one values for that variable, *i.e.* $\log_e(e) = 1$ and $\log_e(1) = 0$. In the immediate example, however this variable is not tested by the Box-Cox transformation.
[18] The discussion of the bootstrap is taken from Efron (1981).
[19] It should be noted that while Efron indicates that 1,000 bootstrap replications will suffice for deriving unbiased predictions and unbiased confidence intervals, and that 50 to 200 bootstraps for unbiased standard errors are necessary, he presents numerical "proof" for the latter assertion with 200 "trials" of $B = 200$ being done. A "trial" represents one set of bootstrap replications. This is perhaps the reason that Efron warns:

> "From a traditional point of view, all of the methods discussed here are prodigious computational spendthrifts. We blithely ask the reader to consider techniques which require the usual statistical calculations to be multiplied a thousand times over. None of this would have been feasible twenty-five years ago, before the era of cheap and fast computation. An important theme of what follows is the substitution of computational power for theoretical analysis. The payoff, of course, is freedom from the constraints of traditional parametric theory, with its overreliance [*sic*] on a small set of standard models for which theoretical solutions are available. In the long run, understanding the limitations of the nonparametric approach should make clearer the virtues of parametric theory, and perhaps suggest useful compromises." (Efron, 1982, pp. 2-3.)

[20] The Box-Cox estimations and the bootstrap derived forecasts and standard errors are accomplished on an 80386 personal computer running OS/2 SE 1.3 in SHAZAM® 6.2—see White, *et al.* (1990). SHAZAM provides an explicit "BOX" command as well as a "DIAGNOS" command with an option, "BOOTSAMP =", that allows the numerical analysis required to obtain unbiased forecasts and their respective standard errors. DIAGNOS pertains to any single-equation estimation routine including the BOX and the ordinary least squares (OLS) estimation. (Having determined the preferred model form through application of the Box-Cox analysis the researcher can define that model as an OLS model.) I want to thank Adrian Pagan who recommended the bootstrap to me for this particular problem and to Ken White for facilitating its implementation.

Table 1

Medicare Capital-Related Cost Model Data Statistics

Variable Name	Mean	Standard Deviation	Minimum	Maximum	Median
MEDKDIS	$473.28	$441.89	$10.932	$12251.	$381.03
MEDRT	41.106	104.71	5.5069	4143.3	29.341
MEDIC	2.4680	28.678	0.10000E-10	1708.0	1.5965
MEDCC	0.30906	0.83520	0.10000E-10	12.586	0.10000E-10
MEDOSC	0.40940	10.751	0.10000E-10	684.38	0.10000E-10
URBAN	0.51395	0.49986	0.00000	1.00000	1.00000
SMHOSP	0.45182	0.49773	0.00000	1.00000	0.00000
MEDEDUC	0.24549	0.43043	0.00000	1.00000	0.00000
TOTBEDS	166.33	157.11	1.00000	1346.0	111.00
OCCPRATE	0.47471	0.20414	0.10041	0.99669	0.46716
DISPROSH	0.14538E+06	0.52773E+06	1.0000	0.97027E+07	1.0000
CMI88	1.1700	0.16464	0.66490	2.1599	1.1613
AGE	18.265	8.7182	1.0005	39.960	18.203

MEDKDIS	Quartiles: 1st	2nd	3rd	4th
mean	$169.3	$304.0	$476.6	$943.4
median	$176.2	$300.8	$470.4	$777.0

Table 2

Log-Log Medicare Capital-Related Cost Model Regression Results

(n = 4265; R^2 in logs = 0.3653; R^2 in levels = 0.1543; $\hat{\sigma}^2_\epsilon$ = 0.53486; LLF = -28,655.5)

[regressand = *ln*(MEDKDIS)]

Regressors Name	Estimated Coefficient	Standard Error	t-Ratio 4252 DF	Partial Corr.	Standardized Coefficient	Level of Collinearity
CONSTANT	5.4026	0.14085	38.358	0.5070	0.00000	n.a.
ln(MEDRT)	0.17203	0.20777E-01	8.2802	0.1260	0.13054	0.1896**
ln(MEDIC)	0.11723E-01	0.97687E-03	12.001	0.1810	0.20117	0.1575**
ln(MEDCC)	-0.23325E-03*	0.97232E-03	-0.23989	-0.0037	-0.36683E-02	0.0601
ln(MEDOSC)	-0.93360E-03*	0.13370E-02	-0.69826	-0.0107	-0.96452E-02	0.0233
URBAN	0.21524	0.21481E-01	10.020	0.1519	0.16048	0.3874**
SMHOSP	-0.18625	0.30233E-01	-6.1605	-0.0941	-0.13827	0.5354**
MEDEDUC	-0.50300E-01	0.23490E-01	-2.1413	-0.0328	-0.32293E-01	0.3645**
ln(TOTBEDS)	-0.14203E-01*	0.21703E-01	-0.65443	-0.0100	-0.19689E-01	0.6522**
ln(OCCPRATE)	-0.84578E-01	0.27500E-01	-3.0756	-0.0471	-0.63437E-01	0.4759**
ln(DISPROSH)	0.28644E-02*	0.18158E-02	1.5775	0.0242	0.21140E-01	0.2232**
ln(CMI88)	1.6620	0.91154E-01	18.233	0.2693	0.33873	0.5179**
ln(AGE)	-0.10746	0.13923E-01	-7.7179	-0.1175	-0.94995E-01	0.0083

* coefficient not statistically different from zero.

** potentially problematic level of collinearity.

Table 3

Medicare Capital-Related Cost Model
Selected Actual and Bias Corrected Predictions and 95% Confidence Intervals
(selected from rank-order by observed values then reordered by predictions)

		mean-value		forecast				mean-value		forecast				mean-value		forecast	
actual	prediction	lower-bound	upper-bound	lower-bound	upper-bound	actual	prediction	lower-bound	upper-bound	lower-bound	upper-bound	actual	prediction	lower-bound	upper-bound	lower-bound	upper-bound
$14	$146	$135	$157	$51	$418	$4,878	$280	$254	$310	$98	$803	$2,391	$529	$488	$574	$185	$1,515
$62	$149	$141	$158	$52	$426	$381	$301	$288	$315	$106	$860	$2,536	$542	$506	$580	$190	$1,550
$41	$172	$156	$189	$60	$493	$30	$304	$286	$323	$107	$870	$4,646	$542	$504	$583	$190	$1,551
$32	$172	$160	$185	$60	$492	$2,619	$309	$295	$323	$108	$882	$380	$558	$532	$585	$195	$1,593
$55	$173	$163	$185	$61	$495	$381	$322	$309	$335	$113	$918	$383	$568	$537	$601	$199	$1,623
$60	$179	$169	$189	$63	$512	$382	$349	$333	$365	$122	$996	$382	$595	$552	$641	$208	$1,701
$380	$180	$169	$190	$63	$513	$2,940	$365	$339	$394	$128	$1,045	$383	$604	$576	$633	$212	$1,725
$11	$198	$189	$208	$69	$566	$52	$366	$345	$388	$128	$1,046	$381	$621	$581	$663	$217	$1,774
$48	$203	$189	$217	$71	$579	$3,659	$377	$359	$397	$132	$1,078	$2,622	$628	$587	$672	$220	$1,795
$56	$208	$198	$219	$73	$594	$381	$394	$378	$412	$138	$1,126	$7,580	$667	$630	$707	$234	$1,906
$380	$209	$201	$218	$73	$597	$380	$410	$389	$433	$144	$1,172	$2,377	$670	$629	$713	$234	$1,915
$52	$210	$200	$220	$73	$599	$10,553	$420	$350	$504	$145	$1,218	$2,881	$682	$649	$716	$239	$1,947
$37	$215	$200	$231	$75	$614	$19	$424	$383	$469	$148	$1,215	$382	$708	$662	$757	$248	$2,025
$61	$219	$210	$229	$77	$626	$44	$429	$406	$453	$150	$1,226	$2,893	$719	$685	$754	$252	$2,052
$60	$223	$204	$244	$78	$639	$380	$462	$434	$491	$162	$1,320	$3,063	$735	$690	$782	$257	$2,101
$44	$230	$219	$241	$80	$656	$3,358	$473	$451	$497	$166	$1,352	$382	$776	$724	$830	$271	$2,217
$53	$230	$215	$246	$80	$657	$384	$489	$450	$531	$171	$1,400	$3,034	$875	$778	$985	$305	$2,513
$380	$230	$219	$242	$81	$658	$12,251	$493	$423	$575	$171	$1,422	$3,603	$969	$883	$1,064	$338	$2,776
$55	$258	$246	$270	$90	$736	$382	$497	$473	$524	$174	$1,421	$2,695	$990	$926	$1,058	$346	$2,830
$383	$270	$246	$297	$94	$774	$381	$504	$470	$540	$176	$1,440	$2,390	$1,129	$1,053	$1,211	$395	$3,229
median																	
$381	$406	$382	$432	$142	$1,160												

Table 4

Medicare Inpatient Capital-Related Pass-Through Costs per Medicare Discharge
Actual and Bias Corrected Predictions and 95% Confidence Intervals
(4265 *ex post* forecasts rank ordered by the predictions.)

			mean-value		estimate		forecast	
	Actual	Prediction	lower-bound	upper-bound	lower-bound	upper-bound	lower-bound	upper-bound
overall								
mean	$473.3	$468.6	$442.3	$496.6	$164.2	$1,336.8	$164.0	$1,339.1
median	$299.6	$452.9	$430.1	$476.9	$158.7	$1,292.0	$158.5	$1,293.6
standard deviation	$441.9	$186.5	$173.1	$201.5	$65.4	$532.1	$65.2	$533.6
1st quartile								
mean	$260.4	$253.9	$241.1	$267.5	$89.0	$724.5	$88.9	$725.5
median	$236.0	$256.4	$244.7	$268.8	$89.9	$731.6	$89.8	$732.4
standard deviation	$240.4	$38.3	$36.7	$40.4	$13.4	$109.4	$13.4	$109.5
2nd quartile								
mean	$400.6	$381.1	$361.3	$402.0	$133.6	$1,087.2	$133.4	$1,088.7
median	$408.8	$380.1	$361.9	$399.2	$133.2	$1,084.4	$133.1	$1,085.6
standard deviation	$407.9	$38.5	$36.7	$41.2	$13.5	$110.0	$13.5	$110.2
3rd quartile								
mean	$525.3	$521.5	$493.1	$551.7	$182.8	$1,487.8	$182.5	$1,490.2
median	$420.7	$522.1	$496.6	$548.9	$183.0	$1,489.4	$182.8	$1,491.2
standard deviation	$490.9	$38.8	$37.6	$41.6	$13.6	$110.8	$13.6	$111.0
4th quartile								
mean	$706.8	$717.9	$673.7	$765.1	$251.6	$2,048.0	$251.1	$2,052.2
median	$1696.7	$681.8	$645.3	$720.3	$239.0	$1,945.0	$238.7	$1,947.8
standard deviation	$456.8	$126.6	$112.0	$143.9	$44.4	$361.1	$44.1	$363.5

Table 5

Real Estate Model Descriptive Statistics
(n = 325)

Name	Mean	Median	Standard Deviation	Minimum	Maximum
SELPRC	$51,762	$48,000	$25,719	$13,100	$280,000
TOTFLR	1416.9	1324.0	591.74	572.00	5276.0
BATHS	1.4108	1.0000	0.61717	1.0000	5.5000
AGE	57.317	55.000	28.381	3.0000	126.00
CONSI	1.0238	1.0000	0.10199	0.75000	1.7500
TOTLST	$44,633	$39,800	$20,739	$11,900	$199,700
STHSIDE	1.6609	n.a.	0.83724	1.0000	2.7183

n.a. — not applicable

Table 6

GLBC, LBT, and LBC Real Estate Model Regression Results
[regressand = $\{SELPRC/\mu_{SELPRC}\}^{(\lambda 0)}$]

Regressor Names	Generalized Linear Box-Cox	Linear Box-Tidwell	Linear Box-Cox	Level of Collinearity
CONSTANT	-0.90421E-01	0.91191	-2.0383	n.a.
TOTFLR/μ_{TOTFLR}	0.15358E-02	-0.20203E-05*	0.91033E-01	0.6944
BATHS/μ_{BATHS}	0.56353E-02*	0.36368E-01	0.40846E-01*	0.5070
AGE/μ_{AGE}	-0.68655E-01	-0.67005E-06	-0.99887E-01	0.5199
CONSI/μ_{CONSI}	1.3079	1.2327	1.4576	0.7256
TOTLST/μ_{TOTLST}	0.55345	0.59443	0.42634	0.8575
STHSIDE/$\mu_{STHSIDE}$	0.67254E-01	0.70085E-01	0.82291E-01	0.2919
λ_0	0.1900	1	0.5700	
λ_{TOTFLR}	5.4056	9.9414	1	
λ_{BATHS}	-11.360	2.0458	1	
λ_{AGE}	0.2583	-5.1069	1	
λ_{CONSI}	0.8857	4.4078	1	
λ_{TOTLST}	0.0498	0.6974	1	
$\lambda_{STHSIDE}$	1	1	1	
Goodness of Fit Statistics				
R^2	0.8804	0.9218	0.9154	
F-statistic	162.998	281.999	427.403	
SEE	0.14366	0.13869	0.14007	
LLF	193.401	180.882	190.398	
$\Theta(q)$	—	25.04(1)	6.01(5)	
$\chi^2_{(q,\alpha=0.01)}$	—	6.63	15.09	

All R^2 values reported are measured on the original, untransformed data.
n.a. - not applicable
* coefficient is not statistical different from zero at 0.95 level of significance.

Table 7

Real Estate Model
Selected Actual and Bias-Corrected Predictions and 95% Forecast Confidence Intervals
(selected from rank-order by observed values then reordered by predictions)

actual	95% CI_F lower-bound	prediction	95% CI_F upper-bound	Naive Transformation Bias: 95% CI_F lower-bound	Naive Transformation Bias: prediction	Naive Transformation Bias: 95% CI_F upper-bound	actual	95% CI_F lower-bound	prediction	95% CI_F upper-bound	Naive Transformation Bias: 95% CI_F lower-bound	Naive Transformation Bias: prediction	Naive Transformation Bias: 95% CI_F upper-bound
$19,000	$9,900	$18,400	$27,600	0.0%	1.1%	-3.3%	$48,500	$32,600	$45,700	$59,000	-0.9%	0.0%	-2.4%
$18,500	$11,100	$21,300	$31,800	-9.9%	0.9%	-0.3%	$46,000	$31,800	$46,000	$59,600	-3.5%	0.7%	-1.2%
$13,100*	$14,300	$22,700	$32,800	7.7%	1.8%	-1.8%	$46,500	$34,100	$46,000	$59,500	3.5%	0.7%	-1.3%
$13,200*	$13,500	$23,200	$33,300	-0.7%	1.7%	-1.8%	$47,000	$32,400	$46,900	$60,700	-4.6%	0.2%	-1.5%
$16,000	$14,900	$24,400	$34,900	0.7%	0.4%	-2.6%	$46,500	$35,000	$47,100	$61,000	3.4%	0.8%	-0.7%
$20,000	$12,300	$24,500	$35,200	-22.0%	0.0%	-2.3%	$46,000	$30,500	$47,400	$60,900	-12.5%	0.2%	-1.8%
$18,500	$15,900	$24,700	$34,700	7.5%	2.4%	-2.3%	$48,900	$32,300	$48,000	$62,200	-6.8%	0.8%	-0.5%
$25,000	$15,600	$24,900	$34,900	2.6%	0.8%	-3.7%	$46,500	$34,100	$49,000	$62,700	-4.4%	0.2%	-1.9%
$20,000	$14,000	$26,300	$37,800	-15.7%	1.1%	0.0%	$48,000	$34,900	$50,500	$64,700	-5.7%	0.4%	-1.2%
$20,100	$15,800	$27,100	$38,700	-7.6%	0.4%	-0.3%	$48,000	$38,200	$52,400	$66,800	0.0%	0.8%	-1.0%
$21,500	$15,500	$27,100	$38,600	-9.7%	0.4%	-0.5%	$95,000	$58,800	$78,900	$97,100	-6.0%	0.0%	-0.2%
$24,100	$18,200	$27,400	$38,200	7.1%	1.8%	-1.6%	$109,900*	$58,200	$79,000	$97,000	-6.0%	-0.1%	-1.4%
$24,900	$18,300	$27,600	$39,200	6.0%	1.1%	0.0%	$96,000	$64,400	$84,500	$101,400	-5.1%	-0.5%	-2.4%
$20,000	$15,800	$28,200	$40,100	-12.7%	1.1%	0.5%	$89,900	$64,200	$86,700	$105,100	-7.8%	0.3%	-0.2%
$48,500*	$19,200	$28,500	$40,200	6.3%	0.7%	-0.7%	$83,300	$71,400	$87,400	$104,800	2.1%	0.2%	-1.2%
$21,500	$18,200	$28,600	$39,900	0.5%	1.4%	-1.0%	$86,000	$79,200	$102,300	$121,500	-5.2%	0.2%	-0.7%
$19,900	$19,000	$32,300	$45,100	-13.2%	-0.6%	-0.4%	$120,500	$84,100	$103,000	$122,100	0.8%	0.7%	-0.5%
$25,000	$21,500	$32,600	$44,900	0.0%	0.9%	-0.2%	$105,000	$82,800	$103,200	$122,300	-1.7%	0.1%	-1.1%
$22,000*	$22,200	$33,000	$44,400	2.7%	1.5%	-1.8%	$98,000	$85,400	$103,600	$122,700	1.1%	0.4%	-0.7%
$24,000	$24,000	$34,500	$46,100	3.8%	0.6%	-2.8%	$150,000*	$80,000	$104,600	$124,500	-6.1%	0.3%	-0.8%
$24,500	$21,900	$34,700	$47,500	-6.8%	0.0%	-0.6%	$109,000	$90,100	$107,600	$127,500	2.4%	0.6%	-0.2%
$47,000	$28,600	$40,400	$53,900	2.4%	0.7%	-0.7%	$95,000	$86,300	$107,800	$126,500	-2.1%	0.6%	-1.2%
$49,500	$25,800	$40,600	$53,700	-9.7%	0.7%	-0.7%	$101,000	$92,100	$109,000	$129,100	3.4%	0.3%	-0.8%
$48,500	$26,800	$41,600	$54,600	-9.3%	0.2%	-1.6%	$126,000	$92,600	$111,200	$131,700	1.1%	0.2%	-0.2%
$49,900	$28,100	$42,500	$56,600	-6.8%	-0.7%	-1.4%	$108,000	$91,500	$112,900	$132,400	-1.9%	0.1%	-1.3%
$49,900	$28,500	$42,800	$55,600	-7.0%	-0.2%	-2.7%	$115,000	$93,500	$117,300	$137,100	-3.3%	0.5%	-0.9%
$47,500	$30,500	$42,900	$56,700	1.3%	0.9%	0.2%	$100,000*	$100,900	$120,900	$142,900	0.9%	0.2%	-0.1%
$48,000	$28,700	$43,500	$57,100	-7.0%	0.9%	-0.5%	$155,000	$130,400	$153,600	$175,900	-0.1%	0.0%	-1.3%
$46,500	$31,500	$44,900	$58,900	-1.0%	0.9%	0.0%	$198,000	$161,300	$187,300	$214,000	-0.1%	0.0%	-0.4%
$47,000	$33,200	$45,600	$59,100	1.2%	0.0%	-1.9%	$280,000	$251,400	$285,000	$320,300	-0.4%	0.1%	0.4%
median													
$48,000	$32,000	$44,400	$58,700	0.9%	0.2%	0.0%							

* actual value falls outside 95% forecast confidence interval.

Table 8

Real Estate Model
Linear Box-Cox Market Analysis Model Regression Results
Bootstrap Predictions and Forecast 95% Confidence Intervals*
(325 *ex post* forecasts rank ordered by the predictions)

	actual selling price	bootstrap: minimum 95% CI_F	bootstrap: mean prediction	bootstrap: maximum 95% CI_F	95% CI_F lower-bound	Naive prediction	95% CI_F upper-bound	Naive Transformation Bias: lower-bound	Naive Transformation Bias: prediction	Naive Transformation Bias: upper-bound	Naive Transformation Bias: lower-bound	Naive Transformation Bias: prediction	Naive Transformation Bias: upper-bound
overall													
mean	$51,762	$37,405	$51,767	$65,916	$38,220	$51,535	$66,630	-$815	$232	-$714	-3.0%	0.5%	-1.2%
median	$47,000	$33,200	$45,600	$59,100	$32,800	$45,600	$60,200	$400	$0	-$1,100	1.2%	0.0%	-1.9%
std. dev.	$25,719	$22,234	$24,798	$27,317	$22,195	$24,805	$27,365	$1,850	$223	$507	6.5%	0.5%	0.8%
1st quartile													
mean	$30,484	$19,835	$31,452	$43,151	$20,547	$31,209	$43,744	-$712	$243	-$594	-4.4%	0.8%	-1.4%
median	$36,500	$21,700	$32,100	$44,100	$21,100	$32,000	$44,600	$600	$100	-$500	2.8%	0.3%	-1.1%
std. dev.	$8,340	$4,151	$4,663	$5,394	$3,871	$4,655	$5,398	$1,628	$186	$466	9.3%	0.6%	1.1%
2nd quartile													
mean	$42,094	$28,307	$41,714	$54,921	$29,238	$41,480	$55,523	-$931	$233	-$602	-3.6%	0.6%	-1.1%
median	$45,000	$26,800	$41,700	$55,700	$28,900	$41,100	$55,100	-$2,100	$600	$600	-7.8%	1.4%	1.1%
std. dev.	$6,110	$2,232	$1,973	$2,281	$1,691	$1,928	$2,254	$1,692	$240	$447	6.2%	0.6%	0.8%
3rd quartile													
mean	$52,997	$36,990	$51,535	$65,842	$37,758	$51,310	$66,617	-$768	$225	-$775	-2.3%	0.4%	-1.2%
median	$54,500	$35,800	$51,100	$65,500	$37,800	$51,400	$66,700	-$2,000	-$300	-$1,200	-5.6%	-0.6%	-1.8%
std. dev.	$5,896	$3,788	$3,839	$4,268	$3,328	$3,818	$4,283	$1,786	$215	$467	5.1%	0.4%	0.7%
4th quartile													
mean	$81,534	$64,541	$82,444	$99,835	$65,404	$82,716	$100,715	-$863	$228	-$880	-1.5%	0.3%	-0.9%
median	$73,900	$55,900	$74,900	$91,500	$58,800	$75,000	$92,900	-$2,900	-$100	-$1,400	-5.2%	-0.1%	-1.5%
std. dev.	$32,937	$28,692	$31,250	$34,015	$28,628	$31,263	$33,886	$2,260	$252	$588	3.9%	0.3%	0.6%

How BLS Projects Employment by Occupation.

Daniel Hecker, Office of Employment Projections, BLS

Every 2 years the Office of Employment Projections in the Bureau of Labor Statistics projects employment growth for more than 500 occupations. The most recent set was published in November 1989 and covers the 1988-2000 period; in November 1991, 1990-2005 projections will be published.

When projecting occupational employment, it is useful to keep in mind that labor is purchased (that is, workers are hired) not because it has any intrinsic value, like cars or restaurant meals, but only because it is useful, when combined with machinery and other factors, in producing things like cars and meals. Economists call this a derived demand—derived from the demand for the products it is used to produce. Over time, the employment of workers in an occupation depends on changes in the level of output of products the occupation is used to produce and on changes in production technology or business practices and organization. These changes affect the amount of labor needed to produce a unit of output (labor productivity) and also the combinations of occupations (occupational staffing patterns) needed.

We project the level of output of goods and services, by industry, using a complex econometric model of the economy, with detailed assumptions about the size of the labor force and the level and composition of GNP. The model is described in detail in the BLS Handbook of Methods.

For each industry, projected levels of output per worker (labor productivity) are applied to industry output projections to get total employment by industry. This paper discusses how we translate these industry employment projections into occupational employment projections.

The primary tool for this translation is the Industry-Occupation matrix, which contains 500 occupations on one axis and 250 industries on the other axis. It covers all wage and salary workers in the economy, about 90% of all employment. (We project self employed workers separately.) The matrix shows, for each industry, how industry employment is distributed across each of the 500 occupations, and the percent each occupation's employment is of the total. The array of percentages is the staffing pattern for the industry, and each percent is known as a "coefficient" or "ratio". There is a set of data for the base year and another for the target year.

Base year industry total employment comes from the Bureau's 790 Establishment Survey and base year occupational staffing patterns from the Bureau's Occupational Employment Statistics (OES) Survey. We project target year industry employment using the complex process just mentioned, project staffing patterns, apply them to target year industry totals, and sum occupational employment in each industry.

We could just mechanically apply base year staffing patterns to target year industry totals and sum to get a target year projection for each occupation. Occupations concentrated in fast growing industries then would grow faster than average and those in slower growing industries, slower.

Our projection process, however, is more analytical than mechanical. At its heart is the interplay of two distinctively different methodologies, and projections are revised and fine-tuned until the projection of each occupation's employment is consistent with both. In the first methodology, we evaluate projections of occupational employment, based on analysis of historical employment trends and research into reasons for these trends and an analysis of their future direction. Based on this we may develop a feel for how fast the occupation should grow (either a specific rate or only a range—about average, much faster than average, or slower than average—for example. In the second methodology, we manually change staffing patterns to reflect our analysis of how technology and othert factors should cause coefficients to change, in selected, or all industries. For example, we could lower coefficients of bookkeepers, who are being affected by computerized recordkeeping, making them grow more slowly than otherwise, and raise coefficients of paralegals as their role in legal work expands. We then examine the resulting growth for reasonableness in terms of each of the methodologies, re-evaluate our analyses, and then revise our projections. With sucessive iterations, we gradually approach projections of occupational employment that are consistent with analysis of trends of both occupational growth and industry staffing patterns.

In cases where we can relate employment to independently projected variables, such as school age population and class size (for teachers) or number of motor vehicles (for mechanics), we project employment outside the system and set target year occupational employment at the independently projected level, forcing coefficients to levels which produce this growth. Even in these cases, we examine the resulting occupational coefficients in each industry for reasonableness and make adjustments where warranted. We may also adjust target year employment if we think an extrapolation of historical time series data is more reasonable than the system-generated figure.

Based on our analysis, we also may conclude that raising or lowering target year employment in an industry where the occupation is concentrated is a more reasonable way to alter employment levels for the occupation. Whenever we change the level of one occupation, we make compensating changes in another, or in all other occupations, to keep each industry's employment at the predetermined level. This system also forces the sum of employment in all occupations to the sum for all industries, set earlier in the process. This is a major advantage over single occupation projections, which may be inconsistent with total future employment growth.

When considering staffing pattern changes, we evaluate both time series data on employment levels and information about reasons for growth. We now have 2 or 3 historical observations of employment from the OES, which surveys industries on a 3 year cycle, plus time series data from the Current Population Survey of households, with industry detail. For some occupations, industry, professional association, or other Federal agency time series data also are available.

For nonquantitative information on trends, we examine Federal agency reports, professional journals, the general press,

and other published sources and we do numerous interviews with business and professional association, academic, and government officials and with practitioners in occupations. Information gathered covers changes in consumption patterns, government regulations, products, and in production processes, including new technologies, business organization, and the structure of work. Some specifics are the impact of computers and other automated equipment; efforts to increase production workers' responsiblities, which reduces the proportion of inspectors and supervisors needed; substitution of materials such as plastics or composites for metals, which affects asemblers and welders; and contracting out of cleaning, food, computer, legal, or other services, which shifts the industry where occupations providing the service are employed. It also includes changes in product mix within industries, say between fast-food, sit-down meals, alcoholic drinks, and entertainment in eating and drinking places, which affects the mix between cooks, counterworkers, waiters, musicians, and bartenders.

In summary, our projection options for an occupation are:

1) Hold coefficients constant—let industry growth drive occupations.
2) Increase or decrease coefficients.
3) Force occupational total developed from independent variables or analysis.
4) Increase or decrease total employment in a key industry, which will change occupational growth, without changing coefficients.

We face a number of problems in making projections, similar to those faced by other forecasters. These include:

1) Data sources may cover a somewhat different population than the one being projected.
2) Historical data often have unexplained year to year fluctuations or may be limited to a few observations, making it difficult to tell what has been happening.
3) There may be no anecdotal evidence to explain turning points or points of sharp change in time series data. Without this, it is hard to tell whether a significant change in recent year data is the beginning of a trend, an aberation, or a statistical error.
4) Short-term swings in the business cycle may obscure long term trends.
5) Data sources may show conflicting trends or may differ from reports from field sources.
6) The inherent difficulty in evaluating the likelihood of past trends continuing or the likelihood of a future technological change and its effect on occupations.

Generally we forecast past trends to continue, unless we have evidence for a change. However, where historic data show very rapid rates of change, we tend to project more moderate change, assuming it can't continue at that pace. The result is that the projected growth rates for individual occupations tend to cluster around the average growth rate more closely than historic rates actually did, leading to inaccurate as well as uninteresting forecasts. Our challenge is to push ourselves to make larger changes in coefficients and growth rates than suggested by our instincts.

In summary, although we have a sophisticated quantitative system for making projections, the output is nevertheless greatly influenced by the analysis of our staff. Also, that we approach occupational growth from two perspectives, iterating until we obtain a projection which is consistent with both, and that we use a system which continually forces us to keep the sum of details consistent with the whole.

Total and Net Occupational Separation Data

Alan Eck, Office of Employment Projections, BLS

Students, counselors, training program planners, personnel specialists, and others need information about projected job openings by occupation—openings resulting from employment growth and the need to replace workers who leave an occupation—to make informed decisions affecting career choices, education policy, and organization recruiting efforts. During the past several decades, information about employment growth has been provided biennually by the BLS employment projections program. While recognizing the importance of replacement needs in estimating job openings, BLS stopped developing estimates of replacement needs in the early 1980's because of concerns about data quality and methods of developing data appropriate for different users.

In 1990, BLS began an extensive research project to review methods used to develop estimates of replacement needs in the past and to determine if improved estimates could be developed. This research summary presents an overview of the results of that research.[1]

Most descriptions of the labor market, such as those based on monthly Current Population Survey data (CPS), are developed from information for a single point in time that provide a snapshot of current conditions. Individuals are classified as employed, unemployed, or not in the labor force. Employed persons are further identified by occupation. For each snapshot taken, whether a month or a year apart, the number of individuals in each category generally does not change very much and, thereby, project an image of great stability in the labor market. However, this is not the case. During any time period, there is a great deal of movement into, out of, and between occupations. Measuring these movements to develop estimates of separations from occupations requires longitudinal data about workers at two points in time, which are much less common than snapshots of current conditions. The research focused on the development of procedures that, using available data, would provide the best measure of movements of workers out of occupations over time.

The research concluded that two distinct types of estimates of occupational separations should be developed to meet the needs of all users. The first, total separations, measures all individuals who leave their occupation. The second, net separations, measures the net of movements of experienced workers into and out of occupations. It was found that both measures of separations are best developed using Current Population Survey data, but through different data elements provided by that survey. Total separations are best measured by identifying the experience of individuals over the period of a year. Net separations are best measured by following age cohorts of workers over a longer period of time. The former finding reinforces research conducted in the late 1970's and early 1980's, whereas the latter results in a new approach in developing net occupational separations.

Prior to discussing the methods of developing estimates of total and net separations, a review of definitions and concepts is presented since a variety of concepts have been used to calculate estimates of occupational replacement needs and job openings over the years. These different concepts result in significantly different estimates of separations for the same occupation that often have confused users of the information. The discussion of the methods of developing the estimates of occupational replacement rates using data from the Current Population Survey is followed by a discussion of the techniques used to apply CPS occupation based replacement rates to the Occupation Employment Statistics (OES) survey based occupational data used in the projections program.

In this research summary, data are presented on total and net separation rates only for selected occupations developed through the research. Data covering all occupations in the BLS projection program are presented in the research report referenced above. The 1992 edition of *Occupational Projections and Training Data* (OPTD) scheduled for publication in spring 1992 also will present comprehensive data, including updated total separation data that will use data from a special supplement to the January 1991 Current Population Survey that was not available when the research was conducted. Results of the research will be incorporated in some of the analyses presented in the 1992-93 edition of the Occupational Outlook Handbook that will be available in spring 1992. Some data are used in the projection articles presented elsewhere in this issue of the Review.

Definitions and concepts

Employment growth. If employment is measured at the beginning and end of a given time period and is observed to increase, that increase is a measure of employment growth. Employment growth, a positive net change in employment, creates opportunities for workers to enter an occupation. It results from increased demand for goods and services in the economy and from changes in the occupational structure of industries and is the source of job openings identified by BLS projections. Note that determining employment growth requires only information about employment at two points in time—no information about separations is required. However, employment growth also may be determined by using information about the labor market dynamics of an occupation. For example, employment growth over a given period also can be calculated by subtracting the number of persons separating from an occupation from the number entering.

Total separations. Total separations identify the flow of individuals leaving an occupation for any reason without regard to persons entering and provide the largest measure of separations. During a given time period, some individuals may leave an occupation for a variety of reasons and must be replaced. Some become employed in a different occupation—the result of a promotion, desire to change careers, loss of existing job, need for a different job while attending school or training, need for a different job while caring for a family, or some other reason. Other occupational leavers stop working because they retire, desire more leisure time or time for an extended vacation, assume family responsibilities, return to school, move out of the geographic area, become ill, or for some other reason. If employment in an occupation is to grow or remain the same, those individuals who left the occupation must be replaced. Thus, total occupational

separations are, in most cases, replacement needs and a source of job openings. However, if employment is declining, occupational separations exceed replacement needs by the employment decline because some persons separating are not replaced. It should be noted that individuals who change employers but remain employed in the same occupation are not included in counts of replacement needs because job changes by these individuals have no impact on the number of openings for persons desiring to enter an occupation.

Net separations. Net separations differ from total separations from an occupation in that they summarize movements of workers into and out of an occupation over a specific time period. Net separations provide an estimate of the number of openings for new entrants to replace workers who leave an occupation.

Employment data, by age, for two points in time are used to estimate net separations. For example, occupational employment, by age, is prepared for a base year and for a second year, 5 years later. Then, changes in employment for each age group in the base year is compared with the 5-year older group in the second year. For example, age group 55-59 in the base year is compared with the group 60-64 in the second year. If employment in the age group increased, a measure of net entrants was observed and the measure of net separations for that age group is zero. *If employment in the age group declined, the decline is recorded as the measure of net separations.* Net separations for the occupation is the sum of net separations for all age groups.

It is important to note that within any age group, individuals may have left the occupation and started working in another occupation, stopped working for any reason, or left the geographic area and are no longer included in employment data. Similarly, individuals entering the occupation may have been working in another occupation, may not have been working, or may have come from another geographic area and are additions to the number of employees. The change measured over the time 5-year period reveals only whether entrants were greater or less than separations, but nothing about the magnitude of total entrants, total separations, or any of their components. The change indicates that the size of the original age group increased or decreased but nothing about the specific individuals comprising it. However, inferences can be made from the age distribution of net separations that explains the movements as illustrated later in this report in a discussion comparing separations for registered nurses and waiters and waitresses.

Replacement Needs. Total job openings consist of employment growth and replacements needs due to total separations. Similarly, net job openings consist of employment growth and replacements needs due to net separations. In developing estimates of replacement needs, the distinction between total and net occupational separations and replacement needs must not be overlooked. When employment in an occupation remains the same or increases over a given time period, replacement needs equal separations. However, when employment declines, replacement needs are less than separations because some individuals leaving an occupation are not replaced.

During a period when employment in an occupation declines, total separations will be greater than if employment increased because more individuals lose their jobs and net separations would be greater not only because more individuals leave, but also because fewer enter the occupation. Since a decline in employment represents individuals who left an occupation and were not replaced, replacement needs are determined by reducing separations by the decline in employment. The section "Projections of separations" discusses the methods used to adjust total and net separation data for employment declines in the period used to develop the replacement rates.

Total job openings Total job openings equals growth plus replacement needs due to total separations and provide the broadest measure of openings in an occupation. Estimates of total job openings are useful for identifying differences in demand for additional employees between occupations. For example, waiters and waitresses, and elementary school teachers employ about the same number of individuals, but total job openings for waiters and waitresses are much higher because annual replacement needs are triple those of elementary school teachers.

Net job openings Net job openings equals growth plus replacement needs due to net separations. For some purposes, total openings estimates are not very helpful. Training program planners, for example, cannot use total job openings estimates to identify the number of teachers to train annually because some openings are filled by teachers who previously left the occupation and by workers employed in other occupations who qualify for teaching positions who do not require additional training. To identify training needs, planners must know the number of openings for new entrants to the occupation that consist of openings due to growth and replacement needs resulting from net separations from an occupation.

Developing measures of total separations

All individuals who leave an occupation—those who transfer to another occupation or stop working for any reason—must be included in a measure of total separations. Producing such a measure requires longitudinal data that includes information about individuals at two points in time. In the late 1970's through the early 1980's using Current Population Survey data, the Bureau of Labor Statistics developed a procedure to estimate the total number of job openings arising from workers who leave their occupation over a one-year period.[2] Annual data are preferable because most data on training program completions are compiled on an annual basis. Thus, annual total separation data facilitate occupational supply/demand analyses.

Complete descriptions of the methodology of developing estimates of total separations and discussions of the limitations of the estimates are presented in the publications referenced above. Briefly, the methodology consists of creating a matched sample over a one-year period from the CPS. Matched data are created for each of 12 months and combined resulting in a sample of about 500,000 persons age 15 and older in the initial year. For the research discussed in this report, matched data for 1986-87 about changes in labor force status then were merged with data on occupational transfers from a special study conducted as part of the January 1987 Current Population Survey, the latest available survey of this nature when the research was conducted. Occupa-

tional transfer data from the January 1987 CPS were used because matched CPS data overstate the number of workers who change occupations. The excessively large estimate of occupational transfers in matched CPS data occurs because individuals may respond differently to the same CPS question about their occupation, responses may be recorded differently by interviewers collecting the data, or recorded information may be interpreted and coded differently by clerks preparing files for computer processing. All these actions result in a different occupation being recorded in the second year when, in fact, none occurred. The results of combining 1986-87 matched CPS data and occupational transfer data from the January 1987 CPS are termed merged data and provide a composite description of movements into, out of, and between occupations over a 1-year period. The procedure results in data which identify the numbers and types of separations and the characteristics of workers who change occupations, become unemployed, or leave the labor force.

Total separation data for occupations with fewer than 50,000 employees in 1986 were judged unreliable because of the limited number of observations in the sample. Data for the remaining occupations were examined individually and if data identifying specific reasons for leaving the occupation appeared suspect, another detailed occupational group was selected to serve as a proxy and provide substitute data.

The CPS is conducted primarily to obtain current data on the labor force status of individuals rather than data which measures changes over time. Therefore, there are significant limitations to the data that describe change. Since the CPS is a household survey that obtains data about persons living at a specific address, one limitation to the matched sample is that information can only be developed from the responses of individuals who do not change residence. Movers tend to change their labor force status more than nonmovers; hence, the separation rates are biased downward because movers are not included. Separation rates also are biased downward because of the exclusion of individuals who die between surveys.

Response and coding errors, however, bias the separation rates upward. For example, if employed persons were incorrectly classified as not in the labor force during the second survey, the matched data would indicate movement where none occurred. Although the net effect of the biases on the movements is not known, the impact of the various limitations are offsetting and not concentrated by occupation.[3]

It must be emphasized that total separation rates developed from merged CPS data are not measured rates based on longitudinal data about individuals but a composite estimate of movements from occupations based on CPS data from two sources. However, the rates are occupation specific and, in addition to their value in estimating job openings, are extremely valuable for describing the labor market. The 1986 edition of *Occupational Projections and Training Data* describes differences between occupations and discusses demographic factors affecting total separation rates.[4]

Developing measures of net separations

Because the classification system used in the Current Population Survey has remained constant since 1983,[5] a comprehensive measure of occupation specific net separations can be developed by using changes in age groups over a 5-year period. When the size of a group increases, a measure of net entrants is recorded; when it declines, net separations are identified. Net changes in an age group capture the net impact of transfers into and out of occupations, immigration and emigration, as well as labor force entrants and separations, including deaths. A 5-year period was chosen so as to reduce the impact of cyclical variations that might accompany a shorter period. However, data for other periods can be developed. Data also can be developed by industry, educational attainment, sex, and a variety of other demographic variables. Thus, this new "cohort" technique provides a powerful tool for analyzing labor market changes.

Employment data for appropriate age groups, by occupation, were developed for 1983-88, 1984-89, and 1985-90. Initially, approximately 850,000 records containing occupation, age, and many other characteristics for all employed persons in all months of 1983 were combined and occupational employment by age group tabulated. The process was repeated to obtain data for desired age groups in 1988. To increase the sample size and reduce cyclical fluctuation, data for the same age groups as 1983 were developed for 1984 and 1985, and data for the age groups used in 1988 were developed for 1989 and 1990. Data on employment by occupation, by age group then were averaged and used to prepare the data used in this report. To simplify the presentation, all references to 1985 data represent averages for 1983, 1984, and 1985; references to 1990 data represent averages for the 1988, 1989, and 1990.

Net leavers in most occupations occur only in the older age groups, generally above age 45. This pattern typically describes individuals leaving in large numbers to retire. A different pattern is displayed in some occupations with the vast majority of all net separations taking place in the youngest age groups. In this case, large numbers of workers probably obtained employment in the occupation when they first entered the work force. When they were ready to begin full-time jobs or when they qualified for higher paying jobs, they transferred to another occupation. In both cases, however, the net separations quantify the number of persons who left the occupation permanently. Table 1, which illustrates how net separations for registered nurses and waiters and waitresses were calculated, also shows these different patterns.

In table 1, employment data by age group for registered nurses and waiters and waitresses in 1985 is compared with a 5-year older group in 1990. For example, the number of registered nurses age 20-24 in 1985 are compared with registered nurses age 25-29 in 1990 and the difference calculated. If the difference is positive, more individuals age 20-24 in 1985 entered than left the occupation. Nothing is known about the magnitude of persons transferring into the occupation, coming from outside the labor force, or coming from another country, or, the magnitude of persons transferring out of the occupation, leaving the labor force, or leaving the country. The difference between the two groups simply identifies the amount by which total entrants exceeds total leavers. If, on the other hand, the difference is negative, more individuals left than entered the occupation. Only a negative difference results in a measure of net separations. Positive

differences are recorded as zero net separations for the age group. The separation rate for an age group is calculated by dividing net separations by 1985 employment in the age group. Net separations for all age groups were totaled and divided by total employment in 1985 to obtain the 5-year net separation rate for the occupation.

Registered nurses experience net separations only in the older age groups. Those age 20-24 in 1985 increased by 104,000 during 1985-90, the largest increase of any age group. Much of the increase probably reflects newly qualified graduates entering the occupation. Because few nurses leave the occupation after becoming qualified, there is no measure of net separations in the younger age groups: The 1985 age group 45-49, is the first age group having a measure of net separations. Most net separations for nurses occur in the 1985 age groups 55-59 and 60-64, ages at which many nurses retire.

A much different pattern exists for waiters and waitresses who experience the largest number of net separations in the 1985 age group age 20-24, and smaller, almost steadily declining numbers in all of the remaining age groups. Only the 1985 age group age 16-19 experiences net entrants. The data suggest that many workers get jobs as waiters and waitresses when they first enter the labor market, or hold part-time jobs while attending school. After gaining experience, completing school, or qualifying for a full-time job, many transfer to other occupations; few remain in the occupation long enough to reach retirement age. Thus, most net separations among waiters and waitresses result from occupational transfers, whereas those among registered nurses are due to retirements. In each case, however, replacement needs exist.

Table 1 also presents information for registered nurses and waiters and waitresses about the percent leavers in each age group. This measure is calculated by dividing net leavers in the age group by 1985 employment in the age group. Information about the percent leavers in each age group is especially valuable because it permits estimates of net leavers in the future. The methodology is discussed in the section "Projections of Separations."

Registered nurses and waiters and waitresses are large occupations and the CPS sample for these occupations provides reliable employment data for each age group. For small occupations such as actuaries, statisticians, and mathematical scientists, n.e.c., shown in the following tabulation, however, the sample is too small and the net separation data are unreliable. For example, statisticians have an irregular distribution of net separations among the age groups and the net separation rate of about 25 percent is inconsistent with rates for other professional occupations.

CPS Occupation (In thousands)	Employed 1985	Net Separations, 1985 actual Age group												
From:		16	20	25	30	35	40	45	50	55	60	65	70	75
To:		19	24	29	34	39	44	49	54	59	64	69	74	99
Mathematical and computer scientists	483	0	0	0	0	0	0	0	0	9	5	2	1	1
Computer systems analysts	293	0	0	0	0	0	0	2	0	3	2	1	0	0
Operations and systems researchers	140	0	0	0	0	0	0	0	0	5	3	0	0	0
Actuaries	13	0	0	1	1	1	0	0	0	0	0	0	0	0
Statisticians	27	0	0	0	2	0	2	0	0	1	1	1	0	0
Mathematical scientists, n.e.c.	8	0	0	0	1	2	0	1	0	0	0	0	0	0

To obtain a separation rate for each detailed CPS occupation, data for another detailed occupation or a summary occupation group were substituted for occupations judged to be unreliable based on examination of the data.

In some cases, a larger detailed occupation that had similar occupational characteristics was chosen as a proxy and the separation and employment data for the proxy occupation substituted for the unreliable occupation data and used to calculate separation rates. In other cases separation and employment data for a summary occupation group was substituted. In this case the procedure was not as straight forward. Note in the preceding text table that for the summary occupation group mathematical and computer scientists, no net separations are measured in the data until age 55. Yet, of the detailed occupations comprising the group, actuaries, statisticians, and mathematical scientists, n.e.c. exhibit net separations prior to that age. The summary occupation, mathematical and computer scientists, does not register those separations because total net entrants in the other detailed occupations—computer systems analysts, and operations and systems analysts—exceeded the total of net separations among actuaries, statisticians, and mathematical scientists, n.e.c. However, to exclude the measure of net separations from the summary occupation would result in an understatement of separations from detailed occupations. To overcome that limitation, net separations in each age group for summary occupations were calculated by summing the net separations for each detailed occupation in that age group. Thus, the net separation data for each age group for the summary occupation group mathematical and computer scientists in the following tabulation is the sum of data measured for computer systems analysts, operations and systems analysts, actuaries, statisticians, and mathematical scientists, n.e.c. Because unrounded data for detailed occupations are used, the totals shown below may not be the sum of data presented in the preceding table.

CPS occupation (in thousands)	Employed 1985	Net separations, 1985 adjusted age group												
		16-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-99
Mathematical and computer scientists	483	0	0	1	3	2	3	2	1	9	5	2	1	1

The adjusted separation data for the summary group mathematical and computer scientists were used to calculate separation rates that became the proxy separation rates for actuaries, statisticians, and mathematical scientists, n.e.c. Net separations for other summary occupations found throughout the occupational structure of the economy were developed in the same manner as mathematical and computer scientists.

Projections of separations

Until this point, all information about separations has described what has occurred in the past. However, the Bureau's Employment Projections Program focuses on future opportunities, a purpose that requires projections of employment change, and in addition, projections of replacement needs due to total and net separations.

Total separations. Total separation rates for all detailed occupations were developed from merged CPS data for the 1986-87 period. As described earlier, total separation rates from proxy occupations were substituted for small occupations because the data appeared unreliable. If employment in the occupation was the same or increased from 1986 to 1987, the 1986-87 total separation rate also was the replacement rate and should be used to estimate replacement needs during a projection period. However, if employment declined, the replacement rate was calculated by subtracting the employment decline from the separations.

To estimate annual average total replacements during a projection period, the 1986-87 replacement rate should be multiplied by the employment in the occupation at the mid-point of the projection period. Although labor market conditions have an impact on the replacement rates, attempts to adjust the rates would be fraught with difficulties because not enough is known about the effect of cyclical factors and other labor market conditions on the rates.

Net separations. To develop a net separation rate for an occupation, changes in employment in a given age group in 1985 were compared with employment in 1990 for a group that was 5 years older. As noted above, data for 1985 actually consist of the average of data for 1983, 1984, and 1985; data for 1990 consist of the average for 1988, 1989, and 1990. If employment for the group increased, no net separations occurred and were recorded as zero. If employment declined, the number was recorded as net separations for that age group. The 5-year net separation rate for that age group was calculated by dividing the number of net separations by employment in 1985. (See table 1.) The 1985-90, 5-year net separation rates for each age group then can be applied to employment in future years to obtain a projection of net separations.

Between 1985 and 1990, employment in most occupations increased or remained the same and the 1985-90 net separation rates, by age, were used without adjustment to estimate replacement needs during the projection period. If employment declined, however, one of several adjustments to the age-specific separation rates was used to obtain a replacement rate that reduced the occupational separation rate by the rate of decline in employment. When the employment decline was less than the number of net separations among persons age 16-49 in 1985, the number of net separations age 16-49 was reduced by the employment decline. The decline was distributed in proportion to the number of net separations in each age group 16-49. This was the most frequently used technique and confines the adjustments to the ages most affected by adverse economic conditions since older workers are more likely to remain employed until they retire. In the remaining cases, the net separations were reduced in a like fashion for persons age 16-54 or age 16-65 depending on distribution of net separations in the occupation and the amount by which employment declined. The adjusted age-specific rates then were used to calculate future net replacement needs for persons employed in 1990.

Table 2 illustrates the method for calculating net leavers over the period, 1990-2005, from the stock of persons employed as registered nurses in 1990. First, net leavers were calculated for 1990-95 by multiplying 1990 employment obtained from the CPS for each age group by the replacement rate in 1985-90 for the same age. Before net leavers in 1995-2000 was calculated, employment in 1995 for each age group was determined by identifying employment in 1990 for a 5-year younger age group and subtracting any net leavers 1990-95. For example, table 2 shows 1995 employment of registered nurses age 55-59 to be 98,000. This estimate of 98,000 was calculated by identifying 1990 employment of nurses age 50-54 (114,000) and subtracting the 16,000 net leavers in 1990-95 from that age group. When employment for each age group for 1995 was developed, it was multiplied by the replacement rate for that age group to estimate net leavers for 1995-2000. The process was repeated to obtain employment for each age group in 1998 and to estimate leavers 2000-2005. Summing the number of net leavers for each of the 5-year groups provided an estimate of net leavers for the 15-year period, 1990-2005. Dividing the net separations for 1990-2005 by 15 yielded annual average net separations; dividing annual average net separations by 1990 employment yielded an annual average net separation rate.[6]

New entrants, individuals who were younger than age 16 in 1990 but can be expected to join the group of employed persons after 1990, are not included in the 1990-2005 estimate of separations. If included, estimates of separations with net transfers in the younger age groups—such as waiters and waitresses—would be larger.

Developing national OES survey based occupational separation rates

The preceding section described procedures for estimating total and net separation annual average rates for detailed CPS occupations for the 15-year period 1990-2005. However, the BLS projections program uses an industry-occupation matrix to estimate current and projected occupational employment data that is based on the Occupational Employment Statistics (OES) survey

occupational classification system. Current and projected OES survey based occupational employment data are used for calculating the employment change component of projected job openings estimates. To obtain the replacement needs components of projected total and net job openings, OES survey based estimates of total and net separations had to be developed. This procedure required total and net separation rates for all detailed OES survey based occupations, including collapsed occupations.[7] Development of OES survey based occupation total and net separation rates was accomplished by identifying the CPS occupation or occupations that are equivalent to the detailed OES survey based occupation and either using the CPS rate directly, or calculating a weighted rate using OES or CPS employment as weights if the occupation consisted of more than one OES or CPS occupations. Table 3 presents data for selected occupations.

TABLE 1. 1985-90 Net separation data, by age group (Numbers in thousands)

1985 Employment Age	1985 Employment Number	1990 Employment Age	1990 Employment Number	Net Change	Net Leavers 1985-90	Percent Leavers 1985-90
Registered Nurses						
16-99	1,398		1,581	182	97	6.9
		16-20	3	1	0	0.0
16-19	1	21-24	88	87	0	0.0
20-24	118	25-29	222	104	0	0.0
25-29	289	30-34	303	14	0	0.0
30-34	263	35-39	290	27	0	0.0
35-39	201	40-44	234	33	0	0.0
40-44	156	45-49	169	13	0	0.0
45-49	123	50-54	114	-10	10	7.9
50-54	112	55-59	96	-15	15	13.8
55-59	76	60-64	44	-32	32	41.6
60-64	44	65-69	13	-31	31	70.2
65-69	10	70-74	3	-7	7	67.4
70-74	3	75-79	1	-2	2	69.5
75-99	1	80-99	0	-1	1	78.7
Waiters and waitresses						
16-99	1,379		1,376	-3	350	25.3
		16-20	327	327	0	0.0
16-19	259	21-24	277	18	0	0.0
20-24	418	25-29	245	-173	173	41.3
25-29	232	30-34	172	-60	60	25.9
30-34	138	35-39	114	-25	25	17.8
35-39	92	40-44	73	-19	19	20.7
40-44	69	45-49	53	-16	16	23.6
45-49	53	50-54	40	-13	13	24.4
50-54	45	55-59	36	-9	9	19.8
55-59	39	60-64	24	-15	15	37.7
60-64	22	65-69	10	-12	12	55.0
65-69	7	70-74	3	-4	4	60.0
70-74	3	75-79	1	-3	3	89.5
75-99	2	80-99	1	-1	1	66.4

NOTE: 1985 data actually are averages of 1983, 1984, and 1985 CPS data; 1990 data are averages for 1988, 1989, and 1990.

Table 2. 1990-2005 Net separation data, by age group
(Number in thousands)

1990 Employment		Percent	Net	1995 Employment		Net	2000 Employment		Net
Age	Number	replaced 1985-90	leavers 1990-95	Age	Number	leavers 1995-2000	Age	Number	leavers 2000-05
Registered nurses									
16-99	1581		112		133			162	
16-19	1	0.0	0	16-19	0	0	16-19	0	0
20-24	90	0.0	0	20-24	1	0	20-24	0	0
25-29	222	0.0	0	25-29	90	0	25-29	1	0
30-34	303	0.0	0	30-34	222	0	30-34	90	0
35-39	290	0.0	0	35-39	303	0	35-39	222	0
40-44	234	0.0	0	40-44	290	0	40-44	303	0
45-49	169	7.9	13	45-49	234	18	45-49	290	23
50-54	114	13.8	16	50-54	156	22	50-54	216	30
55-59	96	41.6	40	55-59	98	41	55-59	134	56
60-64	44	70.2	31	60-64	56	39	60-64	57	40
65-69	13	67.4	9	65-69	13	9	65-69	17	11
70-74	3	69.5	2	70-74	4	3	70-74	4	3
75-99	1	78.7	1	75-99	1	1	75-99	2	1

NOTE: 1985 data actually are averages of 1983, 1984, and 1985 CPS data; 1990 data are averages for 1988, 1989, and 1990.

Footnotes

1 A more comprehensive report on the research is presented in *Total and Net Occupational Separations: A Report on Recent Research*, (Bureau of Labor Statistics, August 1991), which can be obtained upon request from the Bureau of Labor Statistics, Office of Employment Projections, 600 E Street N.W., Washington D.C. 20212. In addition to the research on national data by BLS that is discussed in this research summary, the report included an analysis of the applicability of national data in developing estimates of replacement needs for States. That analysis was conducted by Pat Berkery of the New York State Department of Labor.

2 An early version of the procedure was developed in 1978 and produced 1977-78 data that was presented in *Measuring Labor Force Movements: A New Approach,* Report 581, (Bureau of Labor Statistics, 1980). Modifications to the procedure were incorporated in 1982 and the revised procedure was used to develop 1980-81 data (See Alan Eck, "New occupational data improve replacement estimates", *Monthly Labor Review,* March 1984, pp. 3-10.) The most complete description appears in *Occupational Projections and Training Data,* 1982 edition, Bulletin 2202, (Bureau of Labor Statistics, 1982), appendix B, pp. 67-69. The methodology used to prepare 1986-87 data is identical to that used to prepare 1980-81 data.

3 A more detailed discussion of limitations to merged CPS data appears in *Occupational Projections and Training Data,* Bulletin 2202, (Bureau of Labor Statistics, December 1982). pp. 73-75.

4 *Occupational Projections and Training Data,* Bulletin 2251, (Bureau of Labor Statistics, April 1986). pp. 17-23.

5 In 1983, the CPS converted its occupational classification system to the one used in the 1980 Decennial Census, which was compatible to the 1980 Standard Occupational Classification system. From 1972 through 1982, the CPS used the occupational classification system used in the 1970 census.

6 An annual average net separation rate will vary depending on the projection period because the age distribution changes. Also, separations are based on the number of employed persons in the base period. New entrants to the occupation during the projection period are not included in net separation estimates.

7 The term "collapsed occupations" refers to Occupational Employment Statistics survey based occupations for which national data are not published, but were aggregated in a summary occupation. However, the detailed occupation may be presented in State publications, and therefore, an estimated replacement rate was developed.

The National-Regional Impact Evaluation System II (NRIES II): Improved Structure, Linkages, and Performance of a Multiregional Macroeconomic Model of the United States

C. Thomas Lienesch & John R. Kort, Regional Economic Analysis Division, Bureau of Economic Analysis

I. Introduction

The National-Regional Impact Evaluation System (NRIES) model, developed at the Bureau of Economic Analysis (BEA), was introduced in a 1980 issue of the ***Journal of Regional Science***. In the article, the authors describe NRIES as "... an initial attempt at constructing a model that reflects the unique regional economic and demographic growth patterns which, in turn, determine national growth." The original NRIES was largely a bottom-up multiregional model of the United States. National variable values were the sum of the 51 state values for each variable. State-to-state interaction was modeled by distance-weighted demand variables appearing in the right hand side of endogenous state variable equations. The configuration, based on a gravity model, assumed interstate economic activity was a decreasing function of distance. This article presents the changes and evolution of the NRIES model during the past decade. The general nature of the changes may be summarized by adding to the above quote - "with explicit recognition of the dimensions by which regional and national economies interact."

II. Model Improvements

Since the appearance of the 1980 article, the NRIES model has undergone sufficient evolution to now be labeled NRIES II. The flexibility of the NRIES II model makes it a useful tool for addressing a wide variety of research topics. Many refinements and improvements are direct results of these model applications during the last 10 years. For example, the original labor market configuration of the model was respecified during a research project on immigration policy with the Regional Research Institute at West Virginia University. The labor demand - labor supply linkages were reinforced to more fully capture the impacts of exogenous population increases on state labor markets.

The most comprehensive difference between the original and present version of the model is the addition of interstate interaction indexes based on actual commodity flow data. These indexes for the manufacturing industries replace distance-weighted interaction variables based on the "decreasing with distance" gravity model assumption.

Sectoral detail and coverage is increased with the NRIES II. The model contains over 100 behavioral equations per state compared to 69 in the original model. Gross state product (GSP) equations for the two-digit SIC level manufacturing and services industries and one-digit coverage in the remaining nonmanufacturing industries yield 35 endogenous gross output equations per state compared to 12 in the original model.

In addition to greater sectoral detail, data used in estimating GSP by industry equations are improved. Beginning in 1988, the NRIES II output equations are estimated using BEA's estimates of GSP by industry. These GSP estimates provide a consistent measure of economic activity by industry in each state. The comprehensive and detailed primary data on which BEA's estimates are based are a marked improvement over the previously used earnings data. These primary data allow the BEA estimates to capture interregional economic differences and trends better than estimates based solely on earnings.

Finally, NRIES II expands the "bottom-up" nature of the original model by adopting a "hybrid" approach to regional modelling. In this approach, NRIES II combines both "bottom-up" and "top-down" elements. Variables projected within the national model form the basis of the top-down element. These variables enter individual state models establishing a link by which activity in the national economy directly affects the state economies. Variables projected in the state models form the basis of the bottom-up element. These variables are aggregated to national totals forming national projections of the variables. The sum-of-state variables establish a direct link between individual state economies and the national economy. In this hybrid approach, changes in the individual state economies can both affect, and be affected by, changes in the national economy.

The following section presents the general structure of a typical model in the NRIES II system. Section IV continues the discussion with a more detailed examination of the components and specifications of individual state models. Included is a discussion of the interstate interaction indexes. Section V completes the model description with a discussion of the national model. Means of evaluating the model's performance are presented in Sections VI. The remaining sections discuss some of the strengths and weaknesses of the NRIES II model and provide a summary.

III. Model Overview

The NRIES II is a regional econometric projection and impact model used to estimate the spatial distribution of impacts of alternative policies and to provide annual short- to medium-term projections of state economic activity. NRIES II consists of 51 individual state econometric models, a national model, and a set of interstate interaction indexes reflecting trade flows among states. Regional values are formed by summing the values of variables over the component states.

The following section describes the nature of the NRIES II equation specifications. For illustration, the variables of NRIES II can be divided into the two aggregate categories: endogenous and exogenous. Inclusion in a category depends on whether the value is determined within or outside a particular model. However, in a multiregional system, a variable's inclusion in either of the categories depends on the geographic unit being addressed. A simple example is that output of state and local governments in Idaho is endogenous to the Idaho state model and exogenous to the Montana state model. Therefore, to continue the illustration, each of the two aggregate categories is further divided into intraregional and supraregional. The value of a supraregional endogenous variable is determined outside a specific region model but within the NRIES II model as a whole. For example, the implicit price deflator for construction is a supraregional endogenous variable vis-a-vis any state model because it's determined in the national

model. At the same time, it is an intraregional endogenous variable in view of the national model.

For expository purposes, the underlying structure of a typical regional (state) model can be represented by:

$$(1) \qquad X_{it} = A_i X_{it} + B_i Z_{it} + C_i M_{it} + U_{it}$$

In this framework, the economic and demographic variables (X_{it}) in region i, at time period t, are a function of: intraregional endogenous (X_{it}) variables, intra- and supraregional exogenous (Z_t) variables, and supraregional endogenous (M_t) variables. Region-specific coefficient estimates are represented by A_i,B_i, and C_i with U_{it} representing the error term.

The first term on the right-hand side of (1), $A_i X_{it}$, includes the intraregional endogenous variables. These state-specific variables are endogenously determined within each state model and used as explanatory variables in other state equations. In this way, they provide simultaneous relationships among various sectors of a state's economy. Intraregional endogenous variables include, but are not limited to, GSP and employment by industry, and the components of total personal income.

The second term on the right-hand side of (1), $B_i Z_{it}$, includes all (intra- and supraregional) exogenous variables. Intraregional exogenous variables are determined within NRIES II but "given" to a particular state model. These variables, relatively few in number, have unique values for each state. An example is the level of federal aid to states. More common are supraregional exogenous variables. These variables have the same value for all states. Generally, they are national policy variables determined outside the entire NRIES II model. Examples are the federal corporate profits tax rate, national birth and death rates, the minimum wage, and Social Security benefit parameters.

The third term on the right-hand side of (1), $C_i M_{it}$, includes the supraregional endogenous variables. These variables are endogenously determined within the NRIES II system but outside any single state model. In general, they are endogenous "transmission line" variables providing interaction among the various components of NRIES II. Broadly, they are grouped in three categories depending on the dimension of interaction they provide.

The first broad category contains variables behaviorally estimated at the national level. A "top-down" influence is exerted from the nation to the states when these variables are used in individual state models. Examples are short-and long-term interest rates, consumer prices, and output deflators by industry.

The second category of supraregional endogenous variables are the "bottom-up" component of NRIES II. These variables are formed by summing all individual state values of the variable. Examples are gross domestic product by industry for the nation, total U.S. personal income tax receipts, and U.S. retail sales.

The third category of supraregional endogenous variables provide state-to-state economic interaction. These interstate interaction indexes measure state-by-state demand for a particular state's manufacturing output. The indexes are endogenously determined for 20 two-digit SIC industries. Further discussion of the interstate interaction indexes is presented in the following section on state models.

IV. State Models

Each of the 51 state models contains equations for approximately 320 variables. Over 100 behavioral equations, 21 interstate trade flow indexes, and approximately 200 identities are included. The behavioral equations are estimated using OLS with Cochrane-Orcutt correction for serial correlation when necessary. The coefficient estimates are based on state-level time series data for 1969-1988 for the employment and earnings series and 1963-88 in gross product and most other equations.

Production and Labor Demand

Production

Output, generally considered the most comprehensive measure of economic activity, is estimated by the 35 equations for GSP by industry in constant (1982) dollars. In general, the state GSP equations capture local demand conditions through income or all-industry gross product measures for the state. Output demand from other states is measured by interstate interaction indexes for the manufacturing industries. These indexes are discussed in more detail below. Ratios of state-to-nation real industry wages are included in output equations to capture industrial location changes in response to relative production costs.

Individual industry output equation specifications deviate from the general form through a variety of industry-specific variables. For example, in mining output equations, national mining production indexes at the four-digit SIC level are included. The basic specification for construction output is augmented with the consumer confidence index, various interest rate measures, and an index of growth in construction wages covered by collective bargaining contracts. Reflecting the national market for motor vehicles, state output in motor vehicles is a function of lagged U.S. retail sales of new automobiles and the consumer price index for new automobiles. The specification for the output of state and local governments includes local demand for government services measured by state population and income variables. Real general expenditures by state and local governments are also included as explanatory variables. This reflects the balanced budget restrictions in most states.

Interstate interaction indexes

As stated earlier, the nature of state-to-state economic interaction has fundamentally changed in the NRIES II model. Interstate interaction indexes based on actual commodity flow data replaced the distance-weighted interstate demand variables. Derived from the 1977 Commodity Transportation Survey (CTS), these indexes are calculated for each of the 20 two-digit manufacturing industries. They enter the corresponding two-digit manufacturing output equations as explanatory variables. Data in the CTS provide information on commodity shipments by type, weight, value, mode of transportation, and by origin and destination state for up to five-digit transportation commodity code

classifications. Although the CTS is the most statistically reliable source of information on interstate commodity shipments, only in 1977 is the information considered comprehensive. Consequently, a procedure to augment the 1977 CTS data was developed to construct an annual time series compatible with NRIES II.

First, the following "interaction weight" is constructed from the CTS two-digit interstate trade data:

$$(2) \qquad \frac{(S^e_{ri}-S^m_{ri})}{\sum_{\substack{r=1 \\ r\neq s}}^{51} S^e_{ri}}$$

where:

S^e_{ri} = state s exports of industry i's product to state r

S^m_{ri} = state s imports of industry i's product from state r

The numerator is simply net exports of industry i by state s, with respect to state r. The denominator is total gross exports of industry i from state s, excluding shipments to itself (r not equal to s in the summation). Given the total amount of shipments out of state s, the interaction weight measures the proportion of this total shipped to, and demanded by, state r, on a net basis.

Next, a temporal allocation of the new interaction weight is constructed using relative location quotients (LQ's). The pairwise ratio of output-based LQ's is calculated for each pair of states for each year, 1969-1988, and then put on a 1977 = 1.0 basis as:

$$(3) \qquad \frac{s_i^{LQ} \;/\; r_i^{LQ}}{s_{i,77}^{LQ} \;/\; r_{i,77}^{LQ}}$$

The numerator of the relative LQ in equation (3) is interpreted as state s's export position in industry i, relative to state r's export position. If the ratio is greater than 1 (although both states' LQ may be less than 1), then state s is said to be relatively more self-sufficient in producing industry i's output than is state r. If the ratio is less than 1, than the opposite holds true and state r is relatively more self-sufficient. The numerator, the current period LQ ratio, is divided by the corresponding ratio for 1977 because the export-import positions of states (the interaction weights) are based on the 1977 CTS.

To complete the interaction index, a measure of demand (output or income) from all states r (r not equal to s) is multiplied by the interaction weight times the temporal allocation:

$$(4) \qquad s_i^{I} = \sum_{\substack{r=1 \\ r\neq s}}^{51} \left(\frac{(S^e_{ri}-S^m_{ri})}{\sum_{\substack{r=1 \\ r\neq s}}^{51} S^e_{ri}} \right)^{77} \left(\frac{s_i^{LQ}/r_i^{LQ}}{s_{i,77}^{LQ}/r_{i,77}^{LQ}} \right) r_o$$

In 1977, the interaction index simply equals the interaction weight times output or income because the middle term, the temporal allocation, is equal to one. In all other years, the interaction weight is inflated (deflated) by factors which depend on the extent to which the relative LQ position of s-to-r changed in the current year relative to the 1977 value. If s specializes relatively more in industry i (vis-a-vis r) than in 1977, then state s's interaction weight is inflated (multiplied by a factor greater than 1) above the 1977 value. If s specializes relatively less in industry i (vis-a-vis r) than it did in 1977, then s's interaction weight is deflated (multiplied by a factor less than 1) below the 1977 value.

Employment

The third major group of state variables is comprised of 13 equations for industry employment at the one-digit SIC level. These equations, in conjunction with the equations for gross product, form the labor demand side of state models. In NRIES II, labor demand is derived from the demand for industry output. As such, each employment equation contains the level of output of the corresponding industry. This linear factor demand relationship results in a direct link between the output and employment equations of the model.

The employment-by-industry equations also include ratios of state-to-nation average wage rates capturing the effects of relative labor costs on national industrial location. The inclusion of wages in the employment equations forms one side of an economic-demographic link in the labor market of NRIES II. This link is completed by the inclusion of employment-based variables in the wage equations, discussed below. The government employment equations do not include average wages because employment levels in these sectors generally are not sensitive to wage changes. Government employment equations are generally specified as a function of the corresponding government output.

Demographics and Labor Supply

The supply of labor in each state is modeled by equations for the working age population cohorts and the labor force participation rate. The state civilian labor force is equal to the product of the participation rate and working age population. The difference between the state's supply of labor (working age population times labor force participation) and the state's demand for labor (sum of industry employment, except federal military) yields state unemployment.

A person's decision to enter the labor force is modeled by state labor market characteristics, income variables, and a time trend. The labor market variables capture incentives for members of the state's working-age population in a state to enter the labor force. The time trend captures several effects, including generally increasing real wages and rising labor force participation of women.

Population

Population is modeled for five age cohorts — under 5, 5 to 17, 18 to 44, 45 to 64, and 65 and older. The working age population of a state is defined as the sum of two cohorts — 18 to 44 and 45 to 64. Equations for births, deaths, and net migration complete the demographic coverage of the model. Net migration is determined by an identity equation as the change in total population plus births minus deaths.

A common characteristic of the population-by-age group equations is using lagged values of a younger cohort as an explanatory variable. This captures population aging from one cohort to the next.

Population under five is based on current period births and net migration to the state. Population 5 to 17 is a function of lagged population under 5 and net migration. Net migration is included in the equations for the younger cohorts (under 5 and 5 to 17) to reflect adults migrating with children.

The equations for the 18 to 44 and 45 to 64 cohorts reflect the relatively high mobility of these age groups. State population 18 to 44 is a function of several state-to-nation economic variables capturing the relative attractiveness of a state in the migration decision. In this way, population migrates to states with greater employment opportunities or higher wages relative to the nation. The 45 to 64 year old cohort equations are similar to the 18 to 44 cohort except the coefficient estimates show less emphasis on migration, reflecting decreased mobility with age. Population 65 and over is a function of population 45 to 64 lagged plus a time trend variable.

Births are modeled on the state's endogenous population 18 to 44 and the national average birth rate. Similarly, deaths is a function of the national exogenous death rate applied to the state's endogenous population 65 and over. Wages

The labor market and income are linked by equations for current dollar average annual wages in the one-digit SIC industries. Wages provide the equilibrating force that bridges the labor demand and labor supply sections of the model. State labor market conditions are one of the main determinants of current dollar wages. In turn, state-to-nation wages, in constant 1982 dollars, enter the labor demand equations. In this way, increases in state real wages, relative to the nation, dampen local demand for labor. Conversely, an excess supply of labor in a state will put downward pressure on wage rates.

State labor market conditions and growth in prices are the main determinants of current dollar wages. National price indexes and transformations of various price variables account for the inflation component of earnings growth. Various transformations of state employment and unemployment rates are used to capture the influence of local labor market conditions on average wages.

The product of employment and average annual wages by industry yields total earnings by one-digit SIC industry for each state. In most industries and states, total earnings is the largest component of total personal income.

Income, Consumer Spending, and Taxes

Total personal income by place of residence is the sum of five components: total earnings; dividends, interest, and rents; transfer payments; personal contributions for social insurance; and residence adjustment. Total earnings, the largest of these components, is the product of employment by industry and average earnings by industry. The remaining four components are directly estimated. Dividends, interest and rent is a function of state disposable income, population, and transformations of national interest rates. Transfer payments are modeled through measures of the target populations of the recipients and time trend variables. Personal contributions for social insurance, which are subtracted in calculating total personal income, are based on social security tax rates, covered wages, and employment-based transformations. The residence adjustment which converts personal income by place of employment to personal income by place of residence is based on lags and time trend variables.

Four sources of current dollar retail sales are included in each state model. The retail sales of eating and drinking establishments, automobile dealers, food stores, and other retail establishments are estimated as a function of state demand variables. State population, income, and national price indexes enter the right hand side of these equations. The sum of these four sources yields total retail sales. Total sales is used in state tax equations as an estimate of the sales tax base.

State and local government revenues are divided into own-source revenues and intergovernmental transfers. Own-source revenues are further divided into tax revenues (personal and corporate income and sales taxes) and revenues from miscellaneous sources and user charges (estate taxes, licensing and professional fees). Within the own-source category, tax revenues are specified as a function of variables capturing characteristics of the state tax base. In this way, disposable personal income relates to the personal income tax base; all-industry gross product reflects the corporate related tax base; and total retail sales reflect the sales tax base. Charges and miscellaneous revenues are specified as functions of state income and population variables because these revenues are largely from user charges for services provided to state residents.

Intergovernmental transfers are specified as a function of several variations of a "need index" calculated to measure a state's relative fiscal position in terms of its population and income characteristics. State and local government expenditures are specified as a function of revenues and state unemployment variables. The inclusion of state revenues recognizes the balanced budget restrictions in most states. The unemployment variables provide a measure of cyclical changes in demand for services over the business cycle.

V. National Model

The national model is composed of two major categories of variables largely distinguished by the direction of interaction with other parts of the model. The first category is national bottom-up variables which are derived by summing the individual state values of the variables. The approximately 300 bottom-up variables correspond to all series included in the individual state models except for the interstate trade flow indexes. These sum-of-state variables, based on individual state determined values, allow states to directly affect national economic activity in a classic bottom-up fashion.

The second major category is comprised of national endogenous variables that show little variation across states or, because of data limitations, cannot be estimated for individual states. There are approximately 190 of these variables of which 40 are behavioral and 21 are exogenous policy variables. These variables enter individual state model equations as explanatory variables. This provides a complementary top-down economic influence from the national economy to the state economies.

The endogenous variables of the national model can be further classified into the following four categories: final demand; federal government receipts and expenditures; interest rates; and price indexes and consumer confidence index. Within the final demand sector, gross domestic purchases are the sum of personal consumption expenditures, gross domestic investment, and total government purchases. Personal consumption expenditures for

durable goods and nondurable goods are estimated as a function of state-summed disposable income and nationally determined output prices.

Total government purchases are the sum of state and local government purchases and federal government purchases. The state and local purchases variable is a function of population variables and state and local government output. Federal government purchases are determined as an identity equal to the sum of defense and nondefense purchases. These two variables are national exogenous policy variables determined outside the model.

Total federal receipts are the sum of state-summed personal income taxes, indirect business taxes, corporate profits taxes, and social security contributions. Federal indirect business taxes, primarily excise taxes, are a function of personal consumption expenditures. Corporate profits taxes are a function of: total output; exogenously assumed statutory tax rate; and of labor's share of income expressed as a ratio of wages and salaries to output. Contributions to social security, the remaining component of total federal receipts, is a function of income, employment, and exogenously assumed coverage and benefit rates.

Total federal expenditures, excluding purchases of goods and services, are the sum of federal aid to state and local governments, subsidies less current surplus, net interest paid, and total federal transfer payments. Of these, only transfer payments and net interest paid are behaviorally estimated in the national model. The total transfer payments variable is a function of state-summed transfers to persons, and net interest paid is a function of the short-term interest rate.

Both the long-term and short-term interest rates are endogenous variables. The short-term rate, defined as the yield on six-month commercial paper, is a function of prices and the exogenously assumed money supply. The long-term rate, defined as Moody's commercial bond yield, is a function of the consumer confidence index and short-term interest rate. The consumer confidence index enters as a proxy measure of the risk premium required on longer maturity instruments.

A number of price indexes are estimated as well as the consumer confidence index. There is a price index for each of the industrial categories in NRIES II and for each of the final demand sectors. In addition, indexes are estimated for consumer and producer prices. In general, both output price and consumer price equations emphasize raw material and wage cost explanatory variables.

As the previous discussion shows, the top-down and bottom-up aspects of NRIES II are not independent entities. Linkages are established within the national model by using sum-of-states variables as explanatory variables in many of the equations for top-down variables. These linkages are reinforced by including top-down variables as explanatory variables in many of the individual state equations for sum-of-state variables. This "hybrid" modelling approach, allows individual state economies to both affect, and be affected by, changes in the national economy. This, coupled with direct state-to-state interaction allows NRIES II to simulate a wide range of the economic linkages necessary in analyzing regional economies.

VI. Model Evaluation

Each year the NRIES II equations are respecified to incorporate new and revised data. During the respecification process, the statistical characteristics of each equation and the reasonableness of the coefficient estimates are evaluated. In view of the large number of equations in the NRIES II model, presenting a variable-by-variable evaluation of model performance is prohibitive. However, a general evaluation of the model's performance is provided by analysis of the simulation errors associated with four aggregate variables — all-industry gross product, employment, personal income, and population. A statistic commonly used for error measurement in econometric models is the mean absolute percent error (MAPE). MAPEs are simply the average of the absolute value of the percentage errors for each year.

$$(5) \quad MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t-\hat{Y}_t|}{Y_t} * 100$$

Where n is the number of periods in the simulation period, Y_t is the actual value of variable Y in time period t, and Y-hat$_t$ is the predicted value of variable Y in time period t.

Several cautions should be noted when evaluating MAPEs. MAPEs are sensitive to the length of the historical simulation; simulations over different numbers of years will produce different results. A further caution in using MAPEs is that aggregated variables may mask larger errors at the detailed component level. This also is true for regional aggregation. For example, the average of the individual state MAPEs may be higher than the MAPEs for the U.S.

The MAPEs presented in table 1 are based on a dynamic historical simulation over the 10 year period from 1979 to 1988. This is a difficult period for a model to simulate because it includes a slowing economy in 1979, two recessions (1980, 1982) and one of the longest expansions in the postwar period (1983-88). In spite of mixed economic character of the period, the performance of the NRIES II model compares favorably with that of many regional and multiregional econometric models.

The four columns of table 1 present MAPEs in all-industry gross state product, total civilian employment, total personal income, and total population for the 51 states and U.S.

For the U.S. as a whole, the MAPE is under one percent in each of the four variables. The MAPEs in GSP and total employment are both approximately 0.9 percent. The MAPE for total personal income is somewhat less at 0.49 percent and lowest for population at 0.2 percent.

The average absolute errors for GSP range from 0.56 percent in North Carolina to over 9 percent in West Virginia. The relatively high statistic for West Virginia results from the model overestimating the low rates of output growth in the state's economy during the 1980's. The all-state average of the gross product MAPEs is 2.9 percent.

The average absolute errors in employment range from 0.9 percent in Pennsylvania to 6.4 percent in Washington, DC. The relatively large error for Washington DC reflects the difficulty of modeling the unique character of its federal government dominated, commuter-based labor markets. The all-state average of the employment MAPEs is 2.2 percent.

The average absolute errors for total personal income range from 0.4 percent in Virginia to 6.09 percent in Iowa. The relatively large error in Virginia reflects the model's underestimate of its relatively rapid income growth. The all-state average of the personal income MAPEs is 2.0 percent.

The average errors for population are significantly lower than for the other aggregates. The average errors for total population range from 0.1 percent in Wisconsin to 5.06 percent in New Mexico. The relatively high MAPE for New Mexico results from the model's failure to capture the decreasing growth rates in the state's population over the decade. The all-state average of the population MAPEs is 0.9 percent.

A less structured method of model validation is to frequently compare a particular model's forecasts with "consensus" forecasts. During the last six years, the NRIES II annual baseline forecasts have been compared, on an informal basis, with the projections of other state models. The forecasts used in these comparisons are from single-state models maintained by university and state and local government researchers. Currently, this exchange involves about 100 researchers from all states except Maryland and New York. In the exchange, forecasts are compared each year for a set of common variables. While difficult to generalize from the results of six years of comparisons with 100 forecasts, the NRIES II forecasts generally track well vis-a-vis a particular state's forecast.

Similarly, the NRIES II forecasts of national variables are annually compared with those of the major national macroeconomic models. Again, the NRIES II forecasts of national variables typically are within the ranges set by the national forecasting services.

VII. Advantages and Disadvantages of NRIES II

There are several advantages to the approach employed by NRIES II. First, because of its interregional elements, NRIES II can be used to analyze the regional or spatial distribution of policy impacts. For example, this might include measuring the effects of economic changes in one region on all other regions. Second, NRIES II simultaneously determines the level of both national and regional activity. In contrast, many existing regio..al models distribute given national totals among regions, thereby ignoring the effects that changes in regional activity may have on the Nation as a whole. Third, by integrating regional and national models, the NRIES II structure enables the analyst to ensure that the sum of regional activity is consistent with reasonable forecasts of national activity. In contrast, individual state models, when summed, can overstate or understate growth when no comprehensive national framework is present. Finally, unlike other existing models, NRIES II is able to examine the effects of concurrent national and regional policy changes. This type of application might include the analysis of the effects of construction activity in one state which is funded by a federal tax levied on all states.

As is the case with any econometric model, NRIES II has limitations which affect the ways in which it should be employed. First, because it is an annual model, NRIES II is better suited to projections and impact analysis over a five to eight year timeframe; it is not well suited to analyzing shorter-term cyclical fluctuations. Second, the industrial detail of NRIES II is relatively aggregate; the impacts of changes in economic policy can be analyzed only at the two-digit SIC level for manufacturing industries and for business, health, and legal services and only at the one-digit SIC level for all other industries. Third, NRIES II is not able to provide spatial detail below the state level. In contrast, an I-O framework may be capable of analyzing impacts at the four-digit SIC and county levels. Fourth, NRIES II is predominantly a demand-driven model which, to some extent, ignores supply constraints. These four limitations largely reflect lack of sufficient regional data, rather than any theoretical weakness in the model structure.

VIII. Summary

This paper describes the evolution, over the last decade, of the NRIES II macroeconomic model of the U.S., developed at BEA. The majority of the improvements address the linkages among the various components of the model. Additionally, sectoral detail is expanded for output equations estimated on improved data. The NRIES II model is constantly evaluated and refined. Future work will largely focus on attempting to overcome some of the previously mentioned limitations. In this vein, extending interstate interaction to the nonmanufacturing sectors is one subject of study. Additionally, incorporating a foreign export sector by state and industry in the model is currently being investigated. As new data sources develop attempts are made to integrate the new information into the model structure. Where data are not available, new approaches are tried to generate reasonable proxies of the needed information.

REFERENCES

Ballard, Kenneth, Wendling, Robert M., "The National-Regional Impact Evaluation System: A Spatial Model of U.S. Economic and Demographic Activity," ***Journal of Regional Science***, Vol. 20, No. 2, (1980):143-57.

Isserman, Andrew, Kort, John, "Regional Economic Consequences of U.S. Immigration Policy," Research Paper 9009, Regional Research Institute, West Virginia University (1990).

Kort, John R., "The Specification of Interregional Interaction in a Multiregional Econometric Model: The Use of the Commodity Transportation Survey in NRIES II," U.S. Department of Commerce, Bureau of Economic Analysis, Regional Economic Analysis Division (1984).

Kort, John R., Cartwright, Joseph V., and Beemiller, Richard M., "Linking Regional Economic Models for Policy Analysis," in M. Ray Perryman and James Schmidt, ed., ***Regional Econometric Modeling***, (Boston, MA: Kluwer Academic Press), 1986, pp. 93-124.

Kort, John R., "An Overview of Regional Modeling Methodology and Data Requirements," in ***Frontiers of Business and Economic***

Research Management, R.W. Resek and J. R. Fitch, ed., University of Illinois, 1983, pp. 90-112.

Renshaw, V., Trott, E.A., Friedenberg, H.L. (May 1988), "Gross State Product by Industry, 1963-86," ***Survey of Current Business***, pp.30-46.

SELECTED READINGS

"Annual Economic and Demographic Projections for States, BEA Regions, and the United States," annual report summarizing NRIES II projections.

"The National-Regional Impact Evaluation System II (NRIES II): Improved Structure, Linkages, and Performance of a Multiregional Macroeconomic Model of the United States," paper presented at the Southern Regional Science Association meetings, April 1991, Miami, Florida.

BEA Regional Projections to 2040, Volume 1: States, Bureau of Economic Analysis, U.S. Department of Commerce (Washington DC: U.S. Government Printing Office), June, 1990.

"Regional and State Projections of Income, Employment, and Population to the Year 2000," (with K. Johnson and H. Friedenberg), ***Survey of Current Business***, (Washington, DC: U.S. Department of Commerce, U.S. Government Printing Office), May 1990, pp. 33-54.

"A Comparison of Regional Economic and Demographic Projections from a Medium-Term Macroeconometric Model Before and After Conjoining with a Long-Term Trend-Extrapolation Regional Projections System," paper presented at the Southern Regional Science Association meetings, March 1990, Washington, DC.

"Decomposing Regional growth in Gross state Product (GSP)", BEA discussion paper 41, 1989.

"Regional Economic Consequences of U.S. Immigration Policy," report submitted to the Economic Development Administration, U.S. Department of Commerce, December, 1988.

"An Analysis of Growth Trends in Appalachian States, 1979-92," paper presented at the Southern Regional Science Association meetings, April 1988, Morgantown, West Virginia.

"Estimating the Fiscal Impacts of the Expansion of the Naval Weapons Station, Earle, New Jersey," report submitted to the Office of the Secretary of Defense, U.S. Department of Defense, April, 1987.

"Fiscal Impacts of defense Spending: The Case of Homeporting," paper presented at the Southern Regional Science Association meetings, March 1987, Atlanta, GA.

"Developing a Satellite Regional Fiscal Impact Model," paper presented at the Southern Regional Science Association meetings, March 1987, Atlanta, Georgia.

"Linking Regional Economic Models for Policy Analysis," in M. Ray Perryman and James Schmidt, ed., ***Regional Econometric Modeling***, (Boston, MA: Kluwer Academic Press), 1986, pp. 93-124.

"Estimating the Potential Regional Economic Impacts of Immigration of Latin Americans and Asians: An Application of the National-Regional Impact Evaluation System," report submitted to West Virginia University, May, 1986.

"Modeling the Regional Economic Impacts of Department of Defense Expenditures on Selected Weapons Systems," report submitted to the Office of the Secretary of Defense, U.S. Department of Defense, September, 1985.

"The Specification of Interregional Interaction in a Multiregional Econometric Model: The Use of the Commodity Transportation Survey in NRIES II", paper presented at the Southern Regional Science Association meetings, March 1984, Orlando, Florida.

"Policy Analysis with the National-Regional Impact Evaluation System," paper presented at the North American Regional Science Association meetings, October 1984, Denver, Colorado.

"An Overview of Regional Modeling Methodology and Data Requirements," in ***Frontiers of Business and Economic Research Management***, R.W. Resek and J. R. Fitch, ed., University of Illinois, 1983, pp. 90-112.

Modeling the Regional Economic Impacts of Major New Military Bases, Regional Economic Analysis Division, Bureau of Economic Analysis (Washington, D.C.: U.S. Department of Defense), April, 1983.

"A Multiregional Test of the Conway-Howard Housing Construction Model," ***Journal of Regional Science***, Vol. 23, No. 3, 1983.

"Estimating the Regional Economic Impacts of Solar Energy with Linked Input-Output and Econometric Models," unpublished, U.S. Department of Commerce, 1982.

"A Multiregional Test of the Conway-Howard Regional Housing Construction Model," paper presented at the Southern Regional Science Association meetings, May 1982, Knoxville, Tennessee.

"An Overview of Local Area and Regional Modeling Methodology and Data Requirements," paper presented at the AUBER

conference, October 1982, Knoxville, Tennessee.

"Modeling the Multiregional Economy: Integrating Econometric and Input-Output Models," ***Review of Regional Studies***, Vol. 11, No. 2, 1981.

"Modeling the Multiregional Economy: Integrating Input-Output Models and Econometric Models" (with J. Cartwright), paper presented at the Southern Regional Science Association meetings, April 1981, Arlington, Virginia.

National-Regional Impact Evaluation System; Structure, Performance, and Application of a Bottom-Up Interregional Econometric Model, by Kenneth Ballard, et.al. (Washington, D.C.: U.S. Department of Commerce), 1980.

"Estimating the Economic and Demographic Impacts of Solar Technology Commercialization on U.S. Regions," unpublished, U.S. Department of Commerce, 1980.

"Developing a Spatially Comprehensive Impact Modeling System: A Feasibility Study" (with K. Ballard, J. Cartwright, and R. Gustely), unpublished, U.S. Department of Commerce, 1980.

"Regional Cyclical Sensitivity and Federal Grants: An Analysis of the Allocation and Impact of Countercyclical Aid" (with R. Gustely), paper presented at the Southern Economic Association meetings, November 1980, Washington, D.C.

Table 1. Mean Absolute Percent Errors 1979 to 1988 from NRIES II Dynamic Solution 1979 to 1988

State	GSP	Employment	Income	Population
Alabama	2.51	2.91	2.76	0.92
Alaska	4.45	1.37	1.81	1.62
Arizona	1.63	1.35	0.88	0.88
Arkansas	1.56	1.78	1.58	0.57
California	2.47	2.30	1.88	0.66
Colorado	2.45	2.07	1.25	1.19
Connecticut	2.46	1.86	1.24	0.59
Delaware	1.96	1.89	1.16	0.33
D.C	3.34	6.38	3.60	0.49
Florida	1.31	1.23	2.38	2.43
Georgia	0.94	0.94	0.45	0.72
Hawaii	2.68	3.50	1.38	0.40
Idaho	1.02	1.93	1.71	0.38
Illinois	4.27	2.78	4.53	0.61
Indiana	5.80	5.48	6.09	0.70
Iowa	2.73	1.74	1.46	0.57
Kansas	1.69	1.99	1.60	0.40
Kentucky	3.34	4.90	4.35	2.14
Louisiana	7.31	4.21	3.14	2.60
Maine	3.69	3.06	3.67	1.07
Maryland	1.49	1.14	0.87	0.24
Mass.	2.09	1.27	0.79	0.32
Michigan	1.95	1.57	1.85	0.55
Minnesota	3.68	5.60	5.15	1.55
Mississippi	3.48	0.76	0.86	0.71
Missouri	1.92	1.39	1.43	0.56
Montana	4.95	1.18	1.32	0.77
Nebraska	1.91	2.03	1.75	0.76
Nevada	1.42	2.51	1.78	0.71
New Hamp.	2.41	1.48	1.60	0.88
New Jersey	3.19	2.30	2.49	1.07
New Mexico	2.84	2.00	1.75	5.06
New York	7.45	3.26	3.58	0.79
N. Carolina	0.56	1.32	1.22	0.35
N. Dakota	5.30	1.46	0.91	1.02
Ohio	2.76	2.36	2.28	0.71
Oklahoma	1.75	2.08	1.78	1.04
Oregon	3.63	4.01	4.34	1.98
Pennsylvania	1.12	0.94	0.57	0.60
Rhode Island	1.34	1.72	0.85	0.23
S. Carolina	1.00	1.26	1.06	0.23
S. Dakota	2.05	1.30	0.98	0.61
Tennessee	4.10	3.29	3.25	1.26
Texas	1.38	1.88	1.52	0.87
Utah	1.41	2.78	2.26	1.04
Vermont	5.45	1.03	0.98	0.59
Virginia	2.43	0.89	0.40	0.21
Washington	2.83	3.32	2.90	1.00
W. Virginia	9.23	1.75	1.47	0.83
Wisconsin	0.47	1.02	0.74	0.14
Wyoming	2.55	2.12	2.25	0.91
U.S	0.88	0.93	0.49	0.19

Developing an Effective Forecasting Program: An Economic Approach

Ralph M. Monaco, Economic Research Service

Developing and maintaining a forecasting program for use in policy analysis is an exercise in constrained optimization. Forecast managers allocate available resources to create and maintain a forecasting program that maximizes the production of the forecasting unit. Because there are constraints in producing forecasts — time, for example — there are trade-offs among resource use to maximize the level of production. This paper suggests that explicitly recognizing constraints and their implied tradeoffs can guide the development and maintenance of effective forecasting programs.

What Do Policymakers and Senior Officials Really Want? The first step in solving a constrained optimization problem is to state the objective function. The obvious objective is to make the most accurate forecasts possible. However, in most cases senior officials want more than just a number or a sequence of numbers. They even want more than the most accurate forecasts. Consider the following thought experiment. A policymaker asks her staff for a forecast of the average level of the prime rate for the next calendar year. After a reasonable amount of time a memo returns with the staff's best estimate: 9.2 percent. End of memo. What is the likely reaction? The memo raises more questions than it answers. What monetary policy was assumed? What Federal deficit is consistent with that rate? What inflation rate? How does 9.2 percent compare to the previous several years' rates? What are other forecasters predicting? Why is the staff forecast different (if it is)? The list goes on.

Certainly the thought experiment is simplistic. However, it illustrates that forecasts are only a part, although arguably the most important part, of what forecasters and forecast programs are expected to do for an organization. Most forecasting groups maintain relatively large databases about their area, handle routine data requests, provide analyses — often extensive — of current developments in their area, and conduct analyses of public policy options. The successful "forecasting" group is actually an information support group, and forecast numbers are only one aspect of the information that they are asked to provide.

In this broader context, the underlying objective becomes: maximize the useful information provided about the developments and likely future of [your subject matter here]. It is not entirely clear whether senior managers or policymakers would articulate their goals so broadly. Indeed, there may be a great deal of resistance to accepting as an objective of a "forecasting" group a statement that does not include the word accuracy. Some of this distinction is semantic; more useful information is generally more accurate information — including forecasts. Further senior managers or policymakers may simply not be aware that the stream of "simple" information requests or briefings they require consume a non-trivial amount of resources that are no longer available for further forecast development.

Given the broader role that the typical forecasting group plays in an organization, it is useful to stress that the objectives — either explicit or implicit — are often wider than the presentation of a sequence of numbers. For forecast unit managers, it is often crucial to keep the wider objectives in mind when designing a forecasting program. And it is crucial to understand that the other related tasks that the forecasting unit undertakes will influence the allocation of resources in the unit and affect the size and scope of the forecasting program.

What have we gained by considering the forecasting problem this way? First, by realizing the broader nature of the typical requests on the forecasting unit, we have brought the "other" demands out in the open and made their completion part of the explicit output of the unit. This actually makes the manager's job easier, since it provides a justification to higher-level managers for why resources were devoted to some particular task that is not explicitly generating a forecast. The broader problem also makes clear to those inside the forecast unit that there are other objectives besides the forecast that they must work to attain.

An example might help. Most forecast managers would argue that the data on which their forecasts are based ought to be housed in a database that is well documented to occasional users, simple to use, easily updated, and flexible in its reporting facilities. None of these features is necessary if the sole reason the data are housed is to support a single forecast model. The benefits of a more flexible system accrue when the data are also used for other purposes — often on short notice — perhaps not directly related to generating a new forecast. Devoting resources to the creation or maintenance of a database is difficult to justify both inside and outside the forecasting unit without reference to the broader information task.

The Constraints. Now the set of constraints must be drawn up. Among the constraints are: time to complete the forecast or provide the information, hardware and software available, and personnel — both in terms of numbers of bodies and in what they know or what they can be reasonably expected to learn in the available time. Cost deserves mention alone as an important constraint. These are short-run constraints. In the longer run, hardware and software can be acquired, employees can be trained or new employees hired, and contracts can be let to help with the forecasting process. But cost is extremely important in the long run too. In the current environment of tight agency budgets, it is likely that cost will be the tightest constraint. Faced with increasing demands, and little opportunity to spend more money on forecasting units, some methods for finding hitherto unexploited areas of productivity gains will have to be found. The last section of this paper suggests some ideas for maintaining or improving productivity.

Explicitly considering the constraints facing the forecasting unit is simply a way of confronting the forecast manager with managerial choices. The typical trade-offs that might be faced include:

> Whether to spend more time developing a database or fine-tuning a forecasting model.
>
> Whether to rely on a technique that is simple, easy to update, run, and explain and that leaves more time for answering other questions, or a more complicated forecasting system that requires more maintenance, explanation, and runtime, but leaves little time for answering other questions.

In this framework, the characteristics of the forecasting system itself become choice variables. Less complicated systems are chosen when the cost — both in terms of direct outlays and in the opportunities foregone — of a more complicated system

becomes too high.

The last consideration raises an issue which some might find alarming, namely, that the "quality" of the forecasting model itself can be traded off against some other facet of the "useful information" objective. This notion is anathema to some who believe that there is a "proper" method or model or software to use in forecasting particular variables. These analysts generally believe in the "linear" approach to problem solving. In this approach, the analyst proceeds sequentially through the steps of problem definition, data collection, analysis, making the forecast, writing documentation and preparing the final write-up with. Ideally, the analyst makes only one iteration through the entire process.

In contrast, Morgan and Henrion argue quite forcefully that:

> Without exception, all of the very good analysts we know take ... [a] very different approach... They view the process of analysis as a process of learning and discovery. They let the policy questions and the structure of the problem drive the analysis, but they take neither for granted and they frequently refine or even redefine both. After dealings with large numbers of inexperienced and mediocre analysts, ... we cannot overemphasize the importance of the difference between this iterative conceptualization of the process of policy analysis and the simple linear approach...(p.40)

In the forecasting context, taking the iterative approach to problem solving suggests that we will be constantly refining and rebuilding our models, elaborating in some areas, but simplifying in others. The argument for a "proper" model relies on knowing in advance the exact nature of the forecasting needs and knowing that the needs will not change. This is not usually the case in forecasting.

Constant refinement — elaborating models where the need for further information requires it, but simplifying where needs no longer require it — is one method to keep the productivity of the forecasting unit rising. Generally, models become more complicated as extra variables or sectors are added, while nothing is deleted because it appears to take more work to eliminate variables than to carry them along "automatically." However, these variables or sectors must still be updated and examined, which takes time. Further, when the analysts working most closely with the model leave or move on to other duties, the new analysts are likely to take longer to understand the model if they must waste time on understanding unnecessary linkages.

The Accuracy Issue. Many presume that the issue of proper technique can largely be resolved by looking at forecast accuracy. Here it is important to remember that techniques are not accurate in and of themselves. Rather, the techniques applied to particular data sets result in a model that is more or less accurate. Generalizations about accuracy of techniques are nearly impossible to make given that accuracy of technique is inextricably tied to a particular data set.

Further, even with respect to a range of models designed to predict the same variables, assessing ex-ante forecasting accuracy is impossible (if it were not there would be no need to forecast!) and evidence on ex-post accuracy is often not entirely convincing. This is because all models are imperfect (there are random errors), and, at least in the case of multivariate models, should admit a degree of uncertainty in variable specification and values chosen for exogenous variables. Forecast accuracy comparisons are generally conducted using relatively small samples, and the error bands associated with the forecasts are often wide enough to enclose most reasonable methods. In other words, in the small samples typically used in judging forecast accuracy, randomness (luck) can largely determine which model is judged more "accurate."

In economic forecasting, this problem is compounded by frequent and sizeable data revisions that can alter the accuracy rankings of various approaches. These considerations are hardly comforting to the forecast manager who must decide to commit resources to, say, a complicated and expensive model, in favor of a less complicated and expensive model.

A thornier, and more realistic, problem arises when the forecast contains several variables, and a different model is judged most accurate for each variable. In many cases, especially in multivariate models that rely on the relationships between variables to produce a forecast, a single model will have to be chosen that includes all the variables.

Finally, it should also be noted that pursuing accuracy without reference to the importance of the variable forecasted in its ultimate use can result in a waste of resources. For example, a simple smoothing technique may be able to forecast disposable income for the next year within 5 percent — a not very difficult task — while a more complicated structural econometric model could reduce the error band to 1 percent. The real question is whether the gain in accuracy is worth cost of resources to gain it. As Spivey indicates

> ...but it is the sensitivity of the decision to the analysis that is often of overriding importance, and in this context, engaging in elaborate modelling and delivering more 'precision' than the decision requires can be as bad as too little. (p. 155)

None of the above is intended to argue that forecast accuracy is not an important consideration. It may even be conceded that it is the most important consideration. However, it is not the only consideration and the difficulties of realistically assessing relative model accuracy should give pause to those who might elevate it to paramount importance. Accuracy is seemingly an unambiguous concept that in practice often leads to a great deal of ambiguity.

A final simple case might illustrate that most analysts already think in terms of trading model "quality" for other considerations, but simply do not carry their tendencies very far. Suppose that it was generally known (somehow) that a particular technique applied to a particular data set had been shown in the past to yield much more accurate forecasts than any other technique. We further suppose that it takes about 20 days for that technique to be applied, rechecked, and written up. However, the forecast must be delivered in 10 days. Most analysts would probably find another technique that delivered less accuracy and deliver a forecast, rather than argue that no forecast at all was a preferable solution.

A Small Digression on Model Types. If accuracy is not very helpful in distinguishing among competing forecasting models, what yardstick do we use to choose among alternative techniques and models? Clearly one answer is: one that maximizes the amount of useful information while fitting within the resource constraints faced by the forecasting unit. We can classify models by the constraints under which they must operate. By the way that models are designed or techniques are implemented, they require

different amounts of resources. Some are extremely computer- and data-intensive, for example ARIMA, vector-autoregression and large structural econometric models, while others are not, such as moving averages and smoothing techniques. Some require a great deal of analyst time and judgment — structural econometric models — while others are nearly automatic — simple time-series models.

Where times between forecasts is short, there are few analysts, and perhaps a single personal computer for support, it is clear that the model must function well in a tightly constrained environment. Other environments have looser constraints. Appropriate models are those whose characteristics match their environment, either by design or by fortuitous accident. Ideally, different models would be developed to function well in different environments, even if the subject matter of the forecast was the same.

Taking models developed in one environment and using them in another environment is hazardous. Models developed for tightly constrained environments will seem overly simple and limited in a more resource-rich environment. The result is likely to be a somewhat bored staff or some excess computer capacity. But there are more disastrous results when a model developed in a loosely constrained environment is applied in a tightly constrained environment. Here the result is likely to be missed deadlines, hurried analysis, frustrated personnel, and eventually, a feeling on the part of upper management or the policymakers that the forecasting unit itself is not meeting expectations. Unfortunately, this model migration is common in the Federal government, where models developed in the less constrained university environment have been moved to a more rigid government staff office. Many policy/forecasting units have "dinosaur" models sitting on a shelf that found themselves extinct after being confronted with a poorer resource environment.

The key to avoiding this problem is to design with the constraints in mind in the first place, and to keep a close watch on the changing resource availability. Viewing the forecasting problem in terms of the suggested broader objective and attention to constraints helps to highlight, and presumably alleviate this design problem.

What Can Be Done? If there are many constraints on the forecasting process and the appropriate managerial response accounts for trade-offs implicit in the constraints, what are some strategies that can help do this? Constructing a successful forecasting system is very much like constructing a useful piece of computer software. The literature on the design of large software products is quite helpful in thinking about the construction and maintenance of forecasting systems. Yourdon's "top-down" approach is useful in this context, especially, in combination with an iterative approach to developing forecasting programs. Here are a few thoughts about how to proceed. Not every idea here is useful in every context.

Begin at the end. It is often useful to create a "demonstration" version of what you will ultimately give to a policymaker or senior official. Create a version of the table that you think will be most important. The table cannot be longer than a single page. Be concerned about the table headings, the line labels, the right number of decimals, etc. Fill in the table with made-up, but plausible forecast numbers, using a simple rules like setting forecast growth rates at their average for some historical period. Take account of the identity relationships — like GNP equals the sum of consumer spending and other components. Fill in as much history as you can reasonably type or write. This process can easily be done in a spreadsheet (or even on a piece of paper) over the course of a few days. After this is finished, you have your first forecast, although it may be a long way from your final forecast.

There are several advantages in working backwards. The process of building such a table will organize information and force you to think about interactions among the table lines that you have created. It will highlight where your knowledge and intuition is strongest and where it is weakest. It will highlight what data you need, and perhaps more importantly, what data you have less need for. The advantages will only accrue if the table sticks to the main point, or the most important concept. Adhere strictly to the 1-page limit.

During this process, it is crucial not to "let the perfect be the enemy of the good." You are building a prototype — a framework to begin your analysis. You will revisit this page often during the forecast development phase, and the table will likely be reorganized — possibly many times. But you will have gained the advantage of seeing the forecast developed and displayed in a way that your consumer will see it. It is surprising how often forecasts are developed and decided upon, and then the final table is prepared. Almost invariably, the final table reveals some inconsistency or anomaly. The revelation is a surprise simply because the final table provides a filter on the forecast that had not been used before. Different filters will provide different insights about the forecast. It is important to have looked through the filter that you will be giving to others to view your forecast through before you deliver your results.

Build a Prototype Model. Along with a prototype table, you have just built a prototype model. While some analysts do not think of models as substitutable (economists seem to suffer especially from this problem), pragmatically they are. Prototype models are designed to be simple. Keep them so. Use them to indicate key relationships, but be ready to throw them away when they are too simple. A prototype model also carries the advantage that a forecast is always only a short time away. It may be a very simple forecast, but in a rush-job setting, having a prototype forecasting model can be the difference between meeting a deadline and not.

Keep the Data Lines Short and/or Automated. Try to avoid manually handling data after the first time it enters the forecasting system. Tables, routine information requests, and, ideally, even the forecasts should be easily extractable from a database.

Choose Software that is Flexible and General. Avoid, if you can, using single purpose software, especially for calculating the forecast. These usually require data in special formats and provide output only in a special format. This leads to another data handling task, both on the input and output side. Since estimation packages are generally more limited than, say, spreadsheets, in formatting tables and graphics, this may mean separating the calculation of the forecast values from the estimation of the parameters of the forecast model. That suggestion only makes sense when the model is not re-estimated every time a new forecast is generated.

Have a Good Consistent Forecast Story. Policymakers make policy. That is not an astoundingly astute statement. But it does have some implications about how your work is done and how it is likely to be received. Policymakers, quite naturally, want to have good reasons for why an event is about to happen, at least

sometimes because they want to do something about it. We should not be surprised that policymakers chafe if the reason that some event is likely to happen is that it did this before. (Often that reason is actually translated into a statement like: the forecast is based on the time- series properties of the data, which tend to exhibit a particular pattern.) Now it may be that this is the best that we can do. However, we ought not to be surprised if a policymaker dismisses the forecast because he is looking for a cause and effect relationship — or a way to clarify his thinking about a particular problem. This suggests that cause and effect models may, after all, be more useful to the policymaker than univariate time-series models, including averaging and smoothing techniques.

Use As Many Techniques as You Can. If you are using multivariate techniques, have one main forecasting model, but use a wide variety techniques to provide different data filters. The literature on combining forecasts is large and shows that there are forecast accuracy gains from using several techniques and combining the results with something as simple as averaging. In macroeconomic forecasting there are several somewhat different approaches. Simple univariate autoregressions or other simple time- series techniques, like vector-autoregression, can be reasonably compared to more structural econometric models. This can be taken too far, especially if the competing techniques or models become quite complicated. Using other techniques to supplement a main forecasting model, or as a comparison for a main forecasting model is extremely useful. Even simple univariate techniques that are fairly easy to implement and maintain can be effective. It is probably better to have a wider array of simple, easy to maintain models to compare against a main forecasting model than a single complicated alternative model.

Keep the Main Forecasting Model Simple. As argued above, why waste resources following variables that are not needed or worrying about their forecast values?

Beware models built in other environments. Having contractors build a model for you can be the most effective way to get a reasonably good model, especially if there has been some difficulty in attracting or keeping quality forecasters. However, as mentioned above, in many cases these models do not meld easily with other tools in the forecasting unit and they may not be designed around the resource constraints facing the unit. Many of the mundane features of the model or the system that is provided become the stumbling blocks, such as how the model is updated, how the results are printed etc. These can prove to be more important to smooth functioning of the forecast unit than the theory on which the model is based. Only very close attention to these matters <u>before</u> the model is delivered can help deal with these difficulties.

Spend Nearly As Much Time Working on Presenting Results as Obtaining Results. Build in a considerable amount of time in the process of planning the forecast schedule for simply writing up the results and figuring out better ways to display them. It is axiomatic that the we perform worst in those aspects of our jobs that we do the least often. In many forecasting units, writing for others — especially nontechnicians — to understand is not done all that often and therefore not done all that well. Clearly communicating the essential features of the forecast and what their implications might be is a time-consuming and difficult task. It will not likely be done well if, say, the results of a six-month long forecasting exercise is written up in the week before it is due. A busy senior manager or policymaker will not have time to sort through a series of tables to understand the forecast. Without a simple and clear write-up, the forecast is often not of much use at all. An additional benefit of getting an early start on communicating what the forecast really says is that you are forced to figure out what the forecast really says, rather than passively looking at tables. As many others have noted, clear communication is clear thinking.

Conclusion. This paper has attempted to show the usefulness of viewing the development and maintenance of a forecasting program through the window of the economist's favorite device, constrained optimization. By explicitly considering what is actually desired, what constraints exist, and sometimes just as important, what constraints do not exist, this view helps to crystallize the issues facing the forecast manager, and can help to make a smoother-running, more effective forecasting effort.

References

Morgan, M. G., and Henrion, M., ***Uncertainty***. Cambridge University Press, 1990.

Spivey, W. A., "The Future of Business and Economic Statistics," ***Proceedings of the ASA Sesquicentennial Invited Paper Session: August 6-10, 1989, August 22-25, 1988.***

Wheelwright, S. C., and Makridakis, S., ***Forecasting Methods for Management, 4th Ed.*** John Wiley & Sons, 1985.

Yourdon, E. ***Managing the Structured Techniques***. Prentice Hall, 1979.

Why Do Forecasters Fail to predict the 'Big' (Unusual) Event?

Herman O.Stekler, National Defense University

Most forecast evaluations emphasize the magnitude of the errors which occurred. In economics it is the percentage error in GNP or some other variable. Meteorological forecast evaluations report on the magnitude of the discrepancies between reported and actual measurements of variables such as temperature, wind velocity, etc.These evaluations of weather forecasts also use quantitative measures to assess the validity of probabilistic forecasts.

However, these are evaluations of the predictions of ordinary or customary events. On the other hand, there are 'big' or unusual events which may not be predictable by the techniques used to forecast the customary events. In fact, there is some evidence that this inability to forecast these 'big' (unusual) events is more pervasive than is customarily recognized. These failures may exist in almost all fields in which professional forecasting occurs.

This paper will present examples of these failures to predict the 'big'(unusual) event from a number of fields-with emphasis on economics. It will provide some explanations for the most obvious failures and will indicate that there are some models that can provide insights about the common characteristics of all these failures.

I. Some Examples of the Failures

A. Economics. The inability to predict cyclical turning points has been recognized as one of the biggest failures of economic forecasting. These changes in the phases of the business cycle signal the beginning of a new state(regime) of the economy. These phases do not have a fixed periodicity and occur several (to many) years apart and are generally not predicted in advance although they may be recognized contemporaneously.

B. Military History. There have been a number of military actions where one of the combatants was completely surprised. The best documented incident in which surprise was achieved despite some intelligence warnings was Pearl Harbor. Other events which achieved strategic surprise include the German invasion of the Soviet Union despite Churchill's veiled warnings of the impending event, the Battle of the Bulge in 1944, the Yom Kippur War in 1973, and the recent invasion of Kuwait.

C. Meteorology. There is some evidence that weather forecasters are frequently surprised by the intensity of some storms,i.e. those that produce heavy rain, strong winds, or heavy accumulations of snow.

D. Political Science. There are at least two recorded examples of the blunders of some political pollsters. The magazine poll of 1936 predicted the victory of Landon while in the election of 1948, Dewey was declared a winner even as the votes were being counted.

In a somewhat different vein, there is now some evidence that statesmen are unaware of the long term relative decline of their nation. For example it is now generally agreed that Britain reached its peak in relative economic power around 1870. Yet it was only at the turn of the century that this was generally recognized and acknowledged. There is now a debate in the U.S. whether this country, too, has been on an unrecognized decline in relative power.

E. Technological Developments. The failure to foresee the impact of new technological developments is legendary. Witness the predictions made in the late 1940s that the world's entire demand for computers would be less than ten, or that TV would not be successful because people would prefer to go to the movies rather than sit in front of a box.

More recent examples were our failures to foresee the impact that semiconductors would have on consumer electronics, computers,etc. Finally, there is the history of the VCR and videodisc where the mass market was definitely not foreseen.

However, the technological forecasting literature is replete with forecasts of 'big' breakthroughs that never or still have not occurred. These include the views that there would be a wide spread adoption of robotics or other advanced automation, that virtually all of our electricity would be generated by nuclear power, etc.

These kinds of predictive errors, i.e. forecasts of 'big' events which do not occur are not as common in other fields. In the field of economic forecasting, they are extremely infrequent. A serious recession, which never occurred, was predicted for the immediate post WWII period. Similarly, there were frequent predictions during 1978 and 1979 that a recession was likely. It never occurred, and when there was a recession in 1980, it was not forecast. If it were possible to explain why these predictions of events which did not materialize were made, then we might be able to obtain insights about the causes of the errors that we are investigating the failure to predict the 'big' event.

II. Why Can Such Failures Occur?

At a later stage we will present some explanations of the reasons why there are some predictions of unusual events which do not occur. Now we focus on some of the reasons that might explain why forecasters and analysts fail to predict some events which do occur.

(1). The data or models that are necessary to estimate such an event either are not available or are inadequate for the task. In some cases, models that are designed specifically for making these kind of predictions might have to be developed.

(2). There may be inappropriate processes by which data or the results obtained from models are interpreted. In other words, the reasoning process may be inadequate.

(3). The forecaster or analyst might have biases or blind spots which would prevent the individual from viewing the data correctly. Alternatively, the analyst might value specific errors in such a way that the failure to predict a 'big' event is an assured outcome.

III. What Can be Done to Avoid Such Failures?

A. Develop Models, Techniques and Data. In the field of economics, it has been recognized that one of the biggest failures has been the inability to predict turning points. Since the process of detecting changes in regimes is different from making quantitative predictions, forecasting methods that are designed exclusively to recognize and detect turning points have been developed. These alternative methods include individual leading series, indexes of leading series, rates of change methods, as well as more sophisticated methods based on these data. Studies have shown that, if we are willing to accept false positive errors, these methods can predict or at least detect the major turning points. Since forecasters still fail to predict turning points despite having these tools at their disposal, additional explanations for these failures need to be presented.

In the field of meteorology, it has been argued that new equipment is required to be able to detect the weather patterns that produce severe weather. I do not know whether this will prove sufficient to eliminate these forecasting errors. There are at least two factors which might explain why Britain's relative economic decline was not understood in the late 19th Century.First, at that time there were inadequate data to track the behavior of the economy. Second, there were no models which explained the factors that contributed to a nation's relative economic decline. In any event, nations currently generate the necessary data, and analysts have adequate models to explain relative economic power. However, as will be explained below, there was one additional factor ideology that could explain the British failure to predict their strategic decline. If such a factor were still operable, today's leaders could still fail to foresee strategic developments.

In the area of political forecasting, this emphasis on new data, models, and procedures has probably eliminated the possibility of egregious errors. The failure to predict Truman's victory in 1948 can be explained by the polling organizations terminating their interviews too early. Now pollsters sample the electorate up to (and, via exit interviews, through) the election. In addition, our knowledge of sampling techniques should prevent a repetition of the 1936 fiasco when one organization used an unrepresentative sample and consequently predicted Roosevelt's defeat.

In the technology area, we have learned why there are mistakes of the other kind, i.e. predicting the commercialization of technologies which are never adopted. These errors stem from emphasizing the projections of entrepreneurs who have a vested interest in seeing the technology adopted. It may also result from failing to project the possible improvements that might occur in the existing technology. However, we cannot explain why the ability to project the possible improvements in existing technology is no better than the capability to estimate which of the new technologies will be adopted.

B. The Interpretation Process. A second explanation for these failures is that part of the reasoning process involved in generating forecasts is inappropriate. One hypothesis is that forecasters use Bayesian procedures in predicting these unusual events. This assumes that forecasters begin with subjective probabilities about the likelihood of these events. As new information becomes available, these prior probabilities are revised in a Bayesian manner.

This hypothesized model of the forecasting process was applied to economic data to determine whether specific cyclical turns should have been predicted. The conclusion was that, given the contemporaneous data, the cyclical peaks of 1957 and 1960 should have been predicted, if the prior probabilities had been greater than zero.

The failure to at least recognize these recessions contemporaneously must be attributed to the zero priors that the forecasters must have had. This means that they did not "expect" the recessions and were "surprised" by their occurrence. This still leaves the forecasters' zero prior probabilities unexplained. In the case of economic turning points, Eckstein had said that cyclical peaks were often associated with credit crunches, and prior to "the crunches, there is no reason to look for the turning point." As a generalization, this suggests that analysts reason by analogy and look for patterns that repeat themselves, rather than attempting to understand the dynamics of the process which might lead to the event that ought to be predicted.

Later analyses of the intelligence data that were available prior to Pearl Harbor have indicated that the information about the likelihood of such an attack could have been inferred, <u>if</u> the analyst had considered this a possibility, i.e. if the analyst had not had zero priors. However, given the mass of data being examined, those clues would not have stood out starkly by themselves. It was necessary to search for them among all the available data.

C. Bias or Ideology. However, an unremitting search for facts to substantiate a particular point of view might on one hand lead to false alarms or be another explanation for the failure to predict unusual events. Ideology often blinds one to the facts and leads to the selection of evidence which is consistent with one's prior views, whether or not they are an accurate reflection of reality. If in the face of overwhelming evidence that suggests that an unusual event is likely to occur, there is a single piece of information which contradicts this evidence, and if the analyst seizes upon this datum, the event will not be predicted. Thus it is very important that forecasters and their principals recognize potential biases. This is an important point, for ideology may have been an important contributing factor in the failure to predict Britain's relative decline. The predominant ideology of the late 19th Century was:that government is best that governs least. No policy change would have occurred even if the decline had been foreseen and consequently there was no incentive for a statesman to look for such an event. There was even an ideological bias against that.

D. Forecasters' Preferences. Finally, there is the possibility that practicing forecasters have asymmetric loss functions. They do not place the same values on the two types of possible errors: (1) failing to predict the event and (2) predicting an event that does not occur. Thus the forecasters' subjective costs may determine whether the event is predicted. Suppose an individual were more concerned with the costs of failing to predict an event rather than with those of "crying wolf". In order to issue a prediction that the event in question will occur, this individual would need less concrete information about the likelihood of the event than would the person who was more concerned with false alarms. Again it is important that both the practitioner and the user of forecasts be aware of these asymmetric costs.

IV. Conclusions.

This paper has attempted to show that the failure to predict 'big' (unusual) events is a problem that affects every field in which forecasting occurs. Some of these events may be unpredictable and we should not be concerned with them. Others can be predicted if the data or models were available. To the extent that the models or data can be developed, we are optimistic that the failure to predict these events can be reduced.

On the other hand, to reduce the failures caused by forecasters' reasoning processes, ideology, or loss functions, we must obtain a better understanding of these factors and eliminate those that produce errors. Unfortunately, we do not understand how individuals interpret data, make judgments or issue forecasts. This is obviously an area in which considerable work needs to be done.

Neural Networks and Exchange Rate Forecasts

David Stallings, Economic Research Service

There are several purposes in writing this paper. First, and most obvious from the title, is an effort to describe a type of associative data processing algorithm which may be unfamiliar to many forecasters. Second, an effort is made to apply this technique, called neural networks or parallel processing, to the problem of exchange rates. Neural networks are not "artificial intelligence" in any sense, any more than, for example, a vector autoregression. A neural network is a way of filtering information, mapping a set of input data to output data.

The use of neural network technology also gives us a chance to examine the way that we, as economists, assume people (that is, noneconomists) look at information. Economists do economic forecasting because we believe there are regularities in human behavior that are part of the human condition. We are, I believe, unique among the social sciences in our willingness to anticipate events. However, our representation of human behavior may, in some instances, compromise our ability to achieve a meaningful forecast. There are nonstatistical, nonprogrammatic ways to examine economic behavior.

Economics and other Social Sciences. The major social sciences, other than economics, include history, psychology, sociology, anthropology, demography, and political science. All are similar in that they study either the human being as an artifact or artifacts left by human beings. None are defined as forward-looking. This is not to say that in many ways they do not possess some predictive power. Psychology, for example, has a strong predictive element about individual behavior in certain circumstances.

Economists, however, are haunted, if not occasionally driven, by the desire to forecast. There are several reasons for this. First, and most obvious, economics is a forward-looking discipline: resources move from lower- to higher-valued uses, entrepreneurs anticipate economic profits so as to shift factors, and consumers constantly seek to increase utility. All require that the future be different from the past. No economic agent is currently in bliss, for bliss lies somewhere in a multitude of tomorrows.

Second, there is value in having better knowledge of future events. This value can be transformed into the morally upright search for knowledge. Also, it can be used to allocate rents more efficiently, presumably at least partly into the pockets of economists. We are, by definition, concerned with efficient resource allocation, so there is no harm, if we believe what we say, in reaping the fruits of our labors. Most economists believe rent dissipation, even into our own wallets, is morally upright — we maintain, uniquely among social scientists, that the search for profits is at least as defensible as the search for knowledge.

A third reason that forecasts are useful to economists is that they validate many of our observations about human behavior. Forecasts also, on more than rare occasions, point to our ignorance. Most economists are brought up with the notion that verification is a critical part of our method. This validation process is also one reason that economists have been more successful at prediction in some other social sciences than educated practitioners of those disciplines. I will cite two selective examples, one each from sociology and political science.

Gary Becker (1976) used the tools of economics to discuss population growth. Two very interesting propositions, both very controversial, but largely verified, arise from his work. First, he stated that the opportunity cost of having children is one primary determinant of birthrates. Thus, wealthier countries have lower rates of population growth: higher wages yield a higher opportunity cost to child-rearing.

The second implication of Becker's, concerning population growth, is that monogamous societies have higher birth rates than those practicing polygamy. The reason is simple: births per woman in monogamous arrangements are higher than in polygamous marriages. The second-order conditions used to verify this result prove that not all economics is sterile (or suitable for children).

My selective reading of history tells me that the declining birthrates in the United States in the 1960's and 1970's would have been better forecast with income as an explanatory variable. It is possible, even (but not likely), that the social security "crises" of the 1970's and early 1980's could have been avoided. This, however, presumes a set of policymakers capable of foresight past the next election.

The problem of policy myopia has been explained convincingly using economics of politics models (Downs 1957), with considerable empirical support. The economic invasion of political science has had far better success than its attempted invasion into sociology or demographics. Budget deficits will continue into the distant future, according to the economics of politics (or public choice school), because there are no incentives (other than moral suasion) to reduce them. The politician who taxes more and spends less is frequently a former politician. Second, new taxes are likely to be indirect and hidden: as corporate taxes, user fees, gasoline taxes, or social security. Tax burdens are also easier to place on future generations, who do not yet vote.

Political science literature is now full of techniques used by economists. Many political scientists explicitly incorporate the self-interest paradigm explicitly in analyzing political institutions. This has proven much more effective, for example, in describing the structure of Congress (Fiorina, 1978) than older "power-centered" models.

One social science, however, has resisted economic imperialism. Psychology has no place for the economist, or at least the rational economic individual. Some of the concepts of rationality ascribed to *homo economicus* are alien to the psychologist's understanding of human behavior. Chief among these are time consistency and transitivity of preferences. There has been, to my knowledge, no verification of these assumptions in the psychological literature (Grether and Plott 1979: 623):

> [The results achieved in the psychological literature] are simply inconsistent with preference theory and have broad implications about research priorities within economics. The inconsistency is deeper than mere lack of transitivity or even stochastic transitivity. It suggests that no optimization principles of any sort lie behind the simplest of human choices and that the uniformities in human choice behavior which lie behind market behavior may result from principles which are of a completely different sort from those generally accepted.

The computational ability of the Becker "family," as well as the technical command now required of the political scientist, may be reasons that other social sciences reject economic explanations. We do not "realistically" describe the way people make decisions.

Economics and Human Behavior. Economists have taken much criticism for the way in which we represent human behavior. The self-interest axiom, for example, is difficult to explain: Few people either can, or are willing to, believe the fact that self-interest is not the same as selfishness. However, economists have compounded this error in understanding by representing individual behavior in a ludicrous manner. The following equation, for example, represents the present value of future wages to a typical worker:[1]

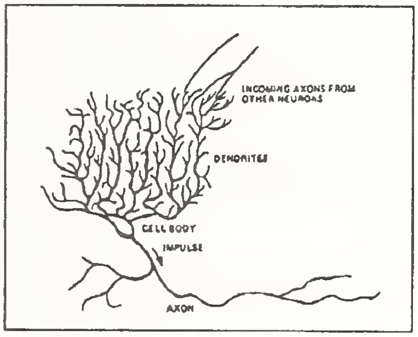
$$w_p = \bar{w}^e \int_0^\infty (1 - e^{-s_0 t}) e^{-(r-g^e)t} dt. \qquad (1)$$

All the individual has to know is calculus, the next wage offer w^e; the nominal interest rate r; the rate of wage inflation g^e; the probability of receiving a wage offer at any time t, s_0; and the probability of having a job at time t, $1 - e^{-st}$. Luckily, however, the typical wage earner, who may not know calculus, only needs to solve the simpler:

$$w_p = \bar{w}^e \left(\frac{1}{r - g^e} - \frac{1}{s_0 + r - g^e} \right). \qquad (2)$$

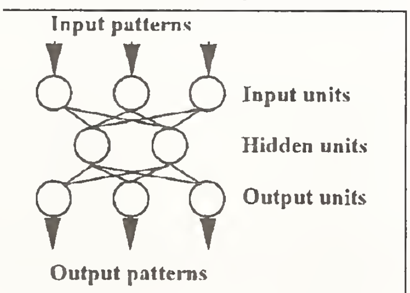
Psychologists, on the other hand, with their allies in biology, have a more realistic way of representing the way that individuals reach decisions (Figure 1). The output area is a long fiber called an axon. The cell, when triggered, sends a signal along the axon to the ends of its fibers. The signal is received by dendrites, the input to another brain cell. An appropriate set of signals (inputs) will trigger the second brain cell to generate an output (along its axon).

Figure 1. Schematic of brain cell



The total number of neurons in the human brain is close to 100 billion (DARPA, 1988). Each is connected to perhaps 10,000 other brain cells. This yields the possibility of 10^{16} interconnections in the human brain. By contrast, a leech has about 1,000. These interconnections can be represented as in figure 2, a typical neural network (Dayhoff, 1990).

Figure 2. Three-layer back-propagation network, fully interconnected



The top layer is an input layer, which corresponds, roughly, to the dendrites. The hidden layer represents neurons, with the output layer acting as axons. The input units receive a pattern, and process it for delivery to each neuron. Each input unit is connected to each hidden unit. The hidden units process each input for delivery to the output units, which generate the output pattern. Weights are assigned at both levels of interconnection (input to hidden and hidden to output), for all connection paths, which represent the strength of each connection.

A neural network is a particular type of input-output model: an input vector, x, is used by the network to produce an output vector y, with a function n:[2]

$$y = n(x) \qquad (3)$$

A simple, and widely used type of network may be represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \qquad (4)$$

The input units (x_1, x_2, x_3) send "signals" to the output unit that are increased or decreased by the factors β_1, β_2, and β_3. This familiar network is often solved via linear regression.

Neural network models are nonlinear models that mimic the brain's architecture. Some applications (White, 1989) include speech decoding, handwriting recognition, and mastering complicated coordination tasks. Neural networks are designed for problems of pattern recognition, classification, nonlinear feature detection, and nonlinear forecasting. Thus far, few economic applications have been found, although the regularities of economic life involve pattern recognition. A neural network model may be appropriate for those "uniformities in human choice behavior which . . . may result from principles which are of a completely different sort from those generally accepted."

A neural network model is estimated recursively. Sample input-output patterns are represented as "learned" behavior. This learning is the result of intensive computation, represented as changing interconnection weights. An input pattern is presented to the input layer, with, for example, one independent variable representing an input unit. Each input is passed, via a weight, to units in a hidden layer, the number of such units being specified *a priori*. Each hidden layer passes a weighted representation of its output to the output layer, which generates an output pattern of fitted values. The output pattern, estimated by the network, is compared to the sample output. The weights are adjusted, beginning with those for the interconnections between the output and hidden layers, then the hidden layers and the input layers. This technique is popularly known as back propagation (or, on occasion, feedforward). The network processes the sample data until a pre-specified error level is reached. Large models may require that the data sample be processed thousands of times before the underlying pattern is "learned."

There are several advantages for the forecaster in using neural network models. First, the fact that the models are nonlinear implies a better fit for the historical data. Second, probability distributions are immaterial. Third, specification is easy; there are no functional forms.

There are also disadvantages to estimation via neural network algorithms. First, because the learning process is computationally intensive, they are time-consuming. A vector-autoregression model that is completed in five minutes may take overnight for back-propagation. Second, the derivation of elasticities or impulse multipliers is very difficult (but not impossible),

although one may interpret interconnection weights in a meaningful fashion. Last, they are poorly understood in the literature. Thus, acceptance of results may be difficult.

Exchange Rates. The problem with the economics of exchange rate determination are straightforward. There is no model that consistently explains long- or short-run exchange rate behavior, because the fundamental determinants are either unknown or affect exchange rates in different ways at different times.

Long-run exchange rates are "supposed" to respond, for example, to long-run changes in aggregate price movements, corrected for economic growth and productivity. This, allegedly, produces a long-run current account balance of zero. Yet, there are widely varying estimates of "equilibrium" exchange rates. Short-run models have been essentially reduced to "speculative bubbles," in which an econometrician tries to guess the univariate time-series model used by the average actor in foreign exchange markets. Surely, we, as economists, can do better.

There are plausible explanations of exchange rate behavior that are well-suited to a "pattern-recognition" system, such as described by neural networks. Clearly, there is some relationship between exchange rates and money. An exchange rate is, for many purposes, the relative price of two currencies. Exchange rates, logically, should also reflect changes in relative prices. Otherwise, arbitrage possibilities would go unexploited. Exchange rates should also manifest relative values for a variety of financial assets denominated in different currencies. Again, arbitrage ensures that this will be the case. Trade balances may be important, but foreign currency transactions are dominated by international financial flows.

The problem with each of these explanations for exchange rate determination is that they all have been correct at certain times, yet incorrect at others. There may well be a threshold, or series of thresholds, below which, say, money supply changes are immaterial. The existence (or nonexistence) of thresholds is one reason to consider a neural network approach to exchange rate modelling.

The architecture of a brain cell, represented in figure 1, is designed to operate in a cost-efficient manner. Not all signals received are strong enough to be processed, nor are all incoming patterns meaningful enough to trigger a response. However, a stimulus or a pattern that exceeds a threshold value will trigger a signal, sent along an axon. Similarly, exceeding an interest rate threshold will, in a neural network model, signal that the exchange rate will change. However, the threshold would depend on the "pattern" of the other input variables, such as prices or money stocks. An econometrician would describe the approach, perhaps, as one with random, or time-varying coefficients.

Empirical Results. A neural network model permits the estimation of multiple outputs. Thus, analogous to a vector-autoregression model, lagged values of dependent values form the set of independent variables. A prediction of exchange rate changes also implies that we must forecast explanatory variables. The model, in essence, generates a recursive set of future values:

$$Y_t = N(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-m}) \qquad (5)$$

$$Y_{t+1} = N(Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-m-1}) \qquad (6)$$

$$Y_{t+2} = N(Y_{t+1}, Y_t, Y_{t-1}, \ldots, Y_{t-m-2}), \qquad (7)$$

and so on.

The set of vector Y's in the example model includes monthly values, 1976-1990, of: yen per dollar; marks per dollar; money market interest rates in Japan, Germany, and the United States; the money supply (M1) in Japan, Germany, and the United States; consumer prices in all three countries; a six- month Eurodollar interest rate; and the size of the Eurocurrency market (assets), measured in dollars. Each Y contains 13 variables, one observation each. All data are from ***International Financial Statistics***.

Lags were arbitrarily set at 1, 3, 6, and 12 months to estimate a learning pattern for the period 1976-1990, for 180 monthly observations. The input pattern thus includes, for each month, 52 variables, and the output pattern 13 variables. The hidden layer was set at sixteen units.[3] Estimation was on an IBM 55sx (16 Mhz 80386sx with math coprocessor) using Neuroshell 2.0. After 12 hours, overnight, output errors were all less than one percent.

Table 1 shows the distribution of weights between the input and hidden units. The first column shows the sum of the lagged weights for each period. The second column, titled "Lag 1" is the sum of the weights going from the input unit to all hidden units for each variable. The most important system- wide variable, Eurocurrency assets, in dollars, was assigned a value of 100, to facilitate comparison.[4] The most important short-run variables are money- market interest rates, consistent with interest arbitrage conditions in foreign exchange markets. The most important longer-run influences are money stocks in Japan, Germany, and Eurocurrency assets, consistent with the monetary approach to exchange rate determination. The U.S. money supply, measured as M1, had little influence.

Fits of the historic data were determined by setting the system error at .005 (fitted vs. actual squared errors for each set of outputs sum to .005). Multiple correlation coefficients (R-squares) for each of the 13 "dependent" variables, not surprisingly, all exceeded .98. Figures 3 through 6 show the historic period for the dollar values of the Japanese yen and German mark, and money market interest rates in those same countries, along with fitted values.

One would suspect, with such robust results for the sample period, that forecasts would prove to be no problem. Forecasts were generated for the first 7 months of 1991. The neural network model predicted the rise in the dollar against both the mark and the yen over 1991, to date, but missed in magnitude (Figures 7 and 8). Nonetheless, the error was less than 3 percent for the yen by August. One-step ahead forecasts for the yen missed by less than one percent, except in March. The error in forecasting the mark was much more severe, reflecting much different conditions in Germany over the forecast period than had been evident during 1976-1990, a result of unification.

The out-of-sample exchange rate estimate errors are partially explained by similar errors in money market interest rate forecasts. Both the German and Japanese money market rates are lower than predicted by the neural network model (Figures 9 and 10). Lower interest rates in Germany and Japan imply a higher value for the dollar. Efficient capital markets, captured in the historic model, also produced a lower Eurodollar rate than forecast (Figure 11).

Reunification conditions in Germany are best reflected in model forecasts for the growth in the German money supply (Figure 12). The forecast money growth was much slower than the 1991 record. As a result, interest rates in Germany were lower than our model indicated. Unification has clearly affected the conduct of monetary policy, and too recently to be "learned" by our model. One could make similar statements about the financial scandals in Japan; the Bank of Japan continues to pump liquidity into the system to alleviate the effects of the shock.

Conclusions

The fact that the exchange rate forecasts were less than perfect might be unsettling to some. However, the robust results and interactions, both long- and short-run, for historical data yield considerable value. "Classic" relationships between money, interest rates, and exchange rates have been confirmed. The increasing importance of Eurocurrency markets in exchange rate determination has also been demonstrated.

There is some promise in the use of pattern-recognition technology to develop economic forecasts. This may be best exemplified by areas in which the statistical tools we currently use hide complex nonlinear relationships among variables. The model estimated provided reasonable results, by historic standards, without incorporating such important variables as GNP growth. Further refinement and extension are clearly in order. Future research will add policy and institutional variables that may prove more useful in expanding and evaluating neural network results.

The use of new technologies also complicates the job of a forecast manager. Which procedure is appropriate for which situation? The proliferation of statistical techniques, models, and model solutions can easily overwhelm those for whom the output is intended. The bottom line here is whether or not neural network procedures add to, or reduce, that confusion. The belief here is that they are important complements to the forecaster's toolkit, to be used in instances suitable for their environment. Only further experimentation can give us an idea of whether they can be used fruitfully or not.

References

Becker, G. (1976). ***The Economics of Human Behavior***. Chicago: The University of Chicago Press.

DARPA Neural Network Study. (1988). Fairfax, Virginia: AFCEA Press.

Dayhoff, J. (1990). ***Neural Network Architectures: An Introduction***. New York: Van Nostrand Reinhold.

Downs, A. (1957). ***An Economic Theory of Democracy***. New York: Harper and Row.

Fiorina, M. (1978). ***Congress: Keystone of the Washington Establishment***. New Haven, Connecticut: Yale University Press.

Grether, D.M. and Plott, C.R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. ***American Economic Review*** 69 (September): 623-638.

Mortensen, D.T. (1970). A Theory of Wage and Employment Dynamics. In E.S. Phelps (ed.) ***Microeconomic Foundations of Employment and Inflation Theory***: 167-211. New York: W.W. Norton and Company, Inc.

Neuroshell. (1988). Frederick, Maryland: Ward Systems Group, Inc.

Rumelhart, D.E. and McClelland, J.L. (1986). ***Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations***. Cambridge, Massachusetts: The MIT Press.

White, H. (1989). Some Asymptotic Results for Learning in Single Hidden- Layer Feedforward Network Models. ***Journal of the American Statistical Association*** 84 (December): 1003-1013.

Footnotes

1 This is not to disparage the work from which this was taken (Mortensen, 1970: 175, n. 12), but is strictly an example of the way economists represent decision-making by a "calculating," rational agent. Any moderately complex utility function would serve to demonstrate the same.

2 Suggested by White (1989).

3 Set as twice the square root of 65 (52 in the input pattern plus 13 in the output pattern), rounded down to the nearest integer. This was suggested in the manual for the software, NeuroShell v. 2.0. The more hidden units, the more accurate the model, but the longer in estimation.

4 Positive and negative weights offset each other for the same input unit, thus comparisons of lagged columns and the total column may be inappropriate.

Figure 3. Japanese yen
Actual vs. fitted values from a neural network

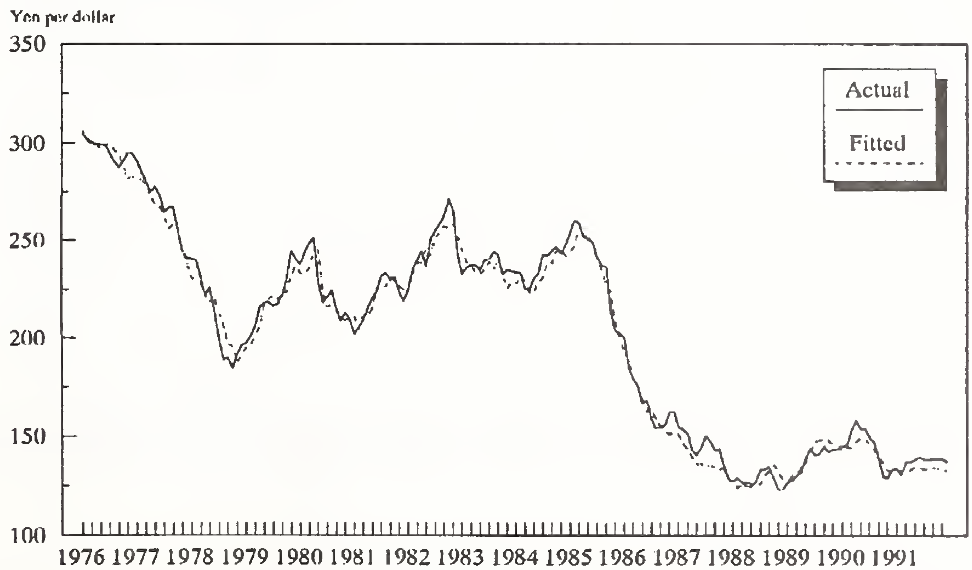


Figure 4. German mark
Actual vs. fitted values from a neural network

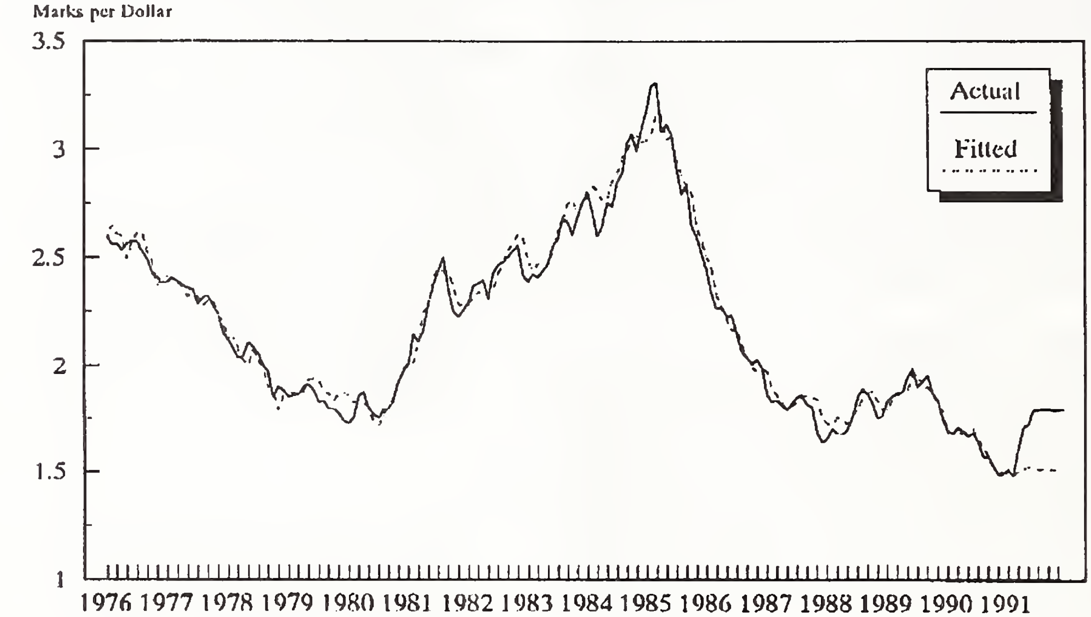


Figure 5. Japan money market interest rates
Actual vs. fitted values from a neural network

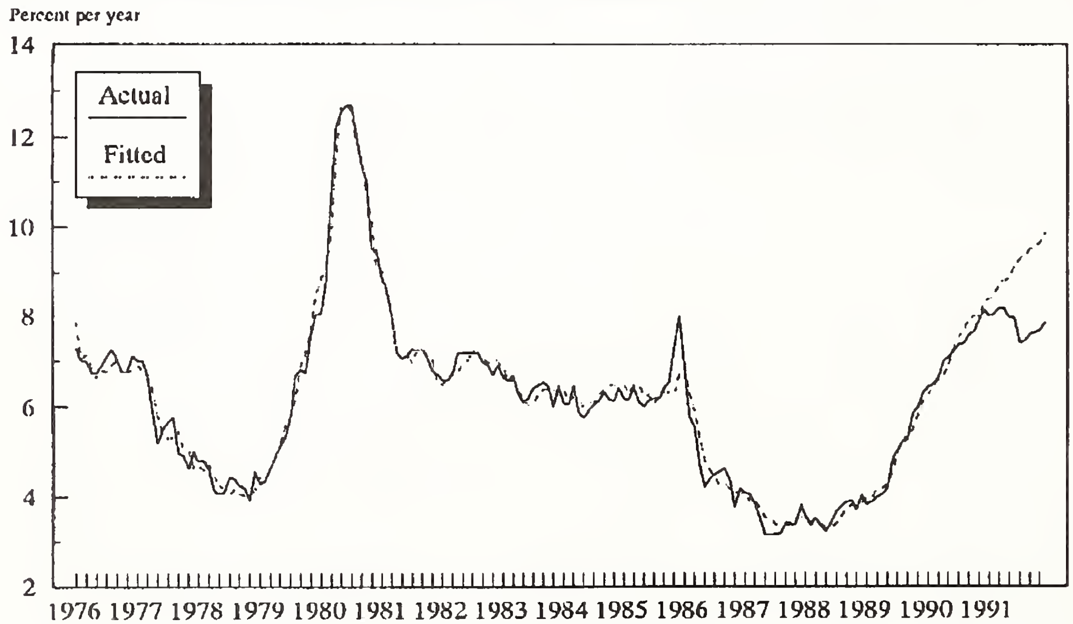


**Figure 6. Germany money market interest rates
Actual vs. fitted values from a neural network**

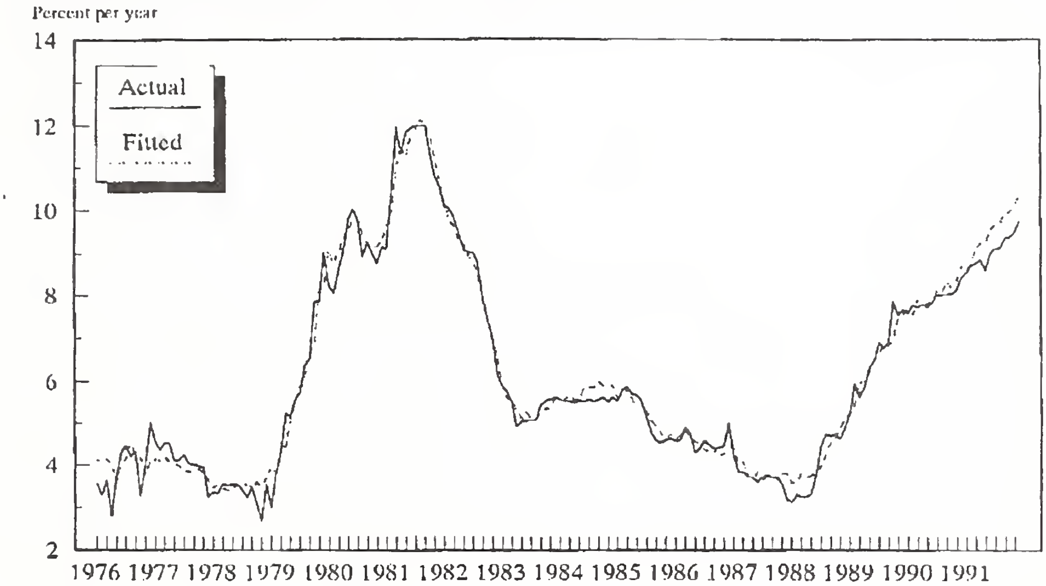


**Figure 7. Forecast vs. actual, 1991
Japanese yen**

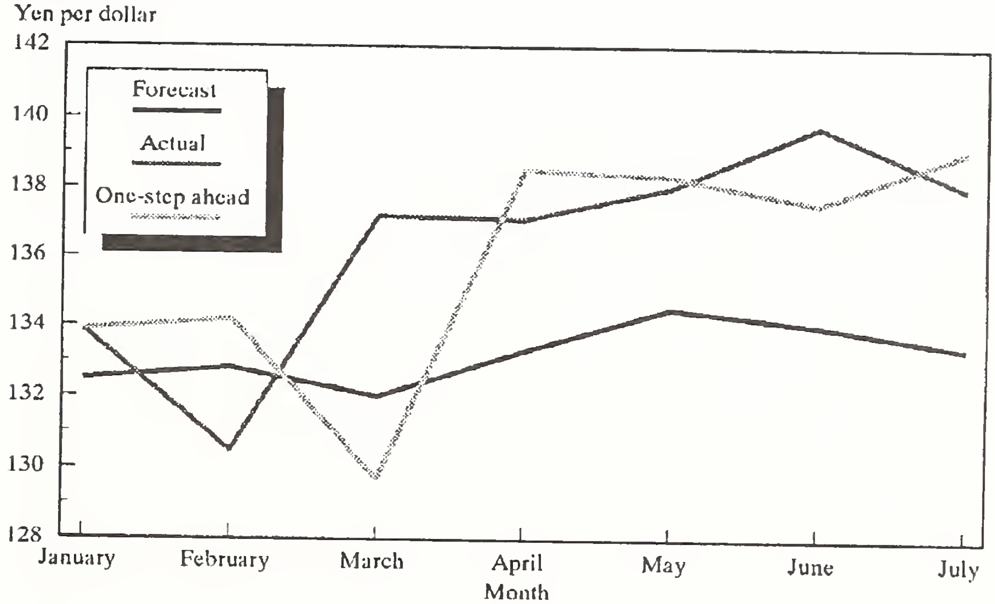


Figure 8. Forecast vs. actual, 1991
German mark

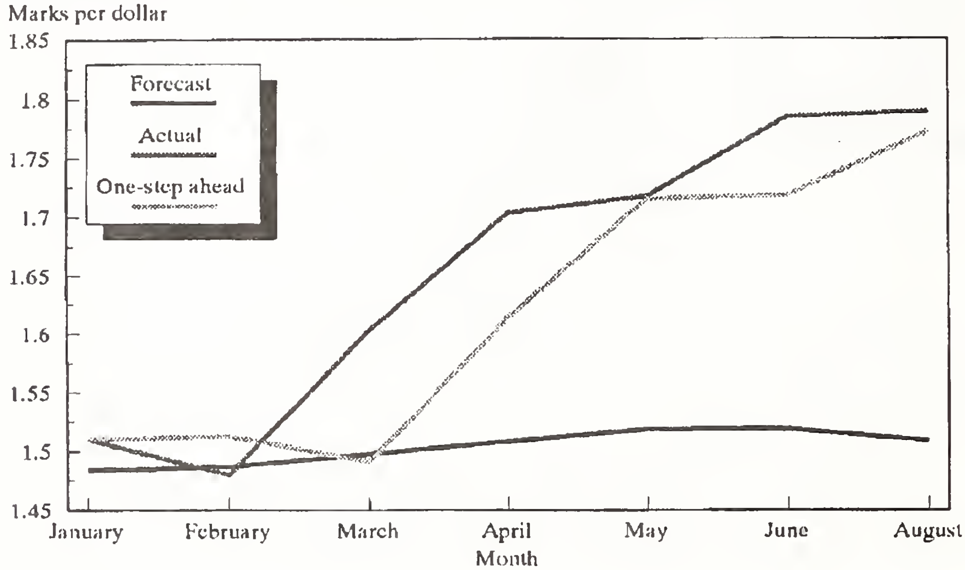


Figure 9. Forecast vs. actual, 1991
German money market rate

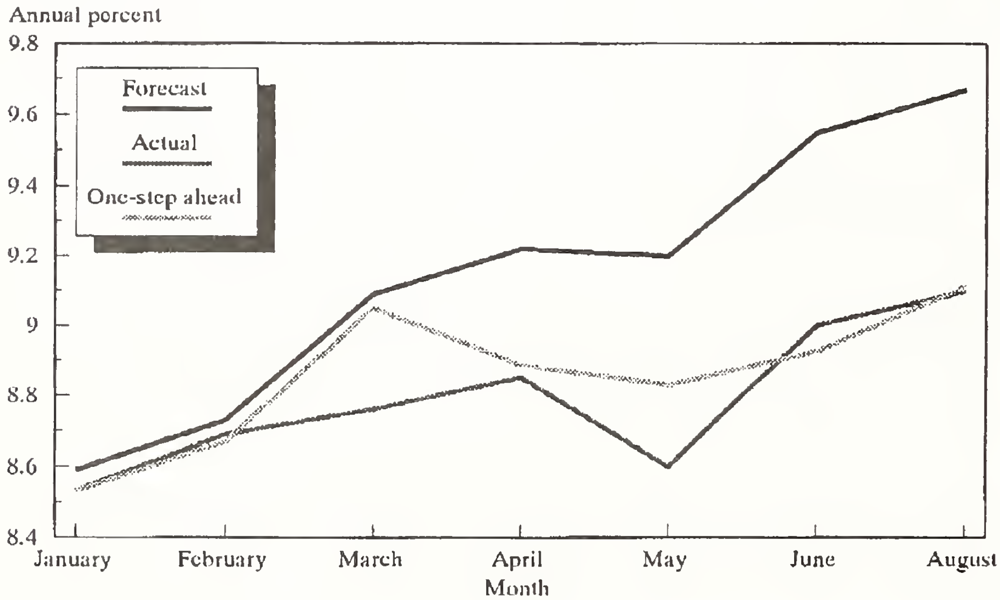


Figure 10. Forecast vs. actual, 1991
Japan money market rate

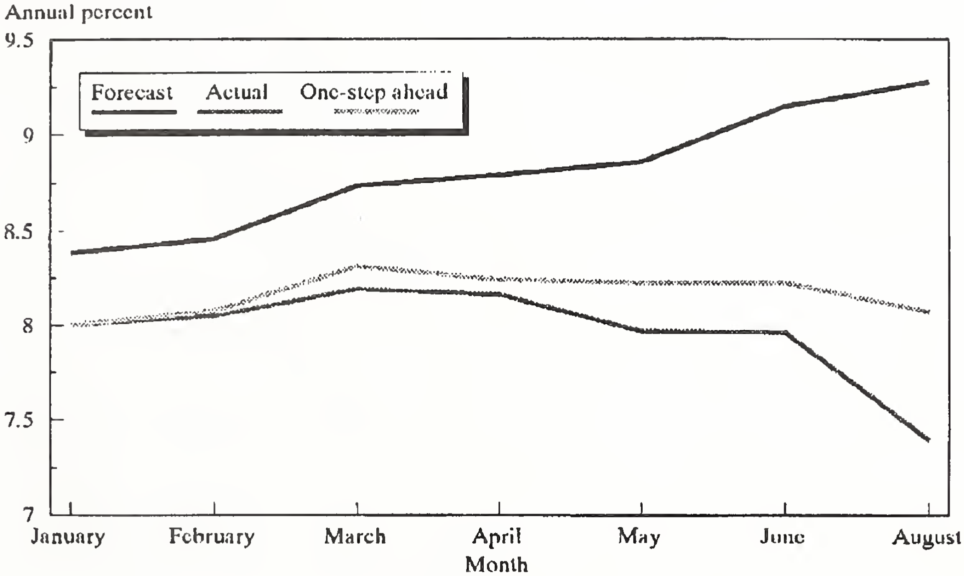


Figure 11. Forecast vs. actual, 1991
6-month Eurodollar rate

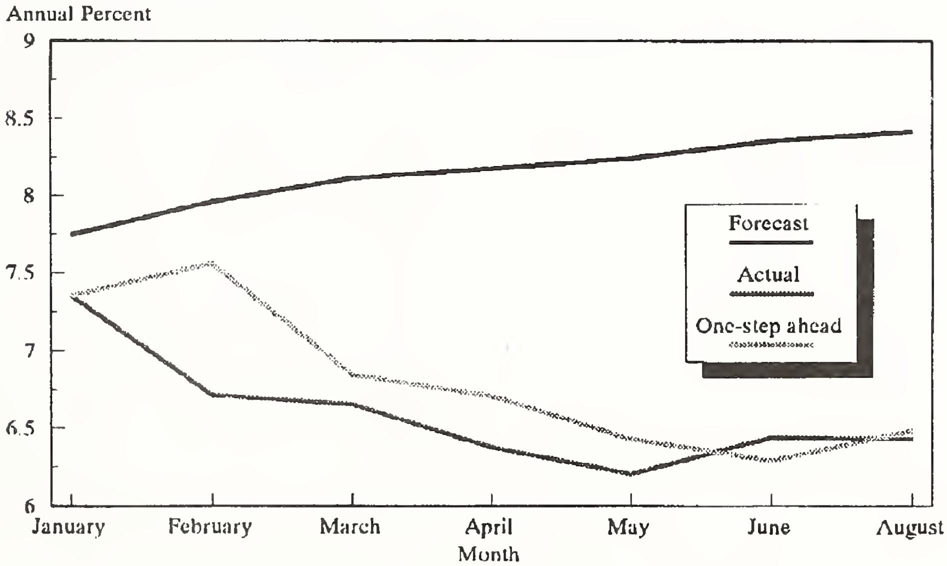


Figure 12. Forecast vs. actual, 1991
Germany M1

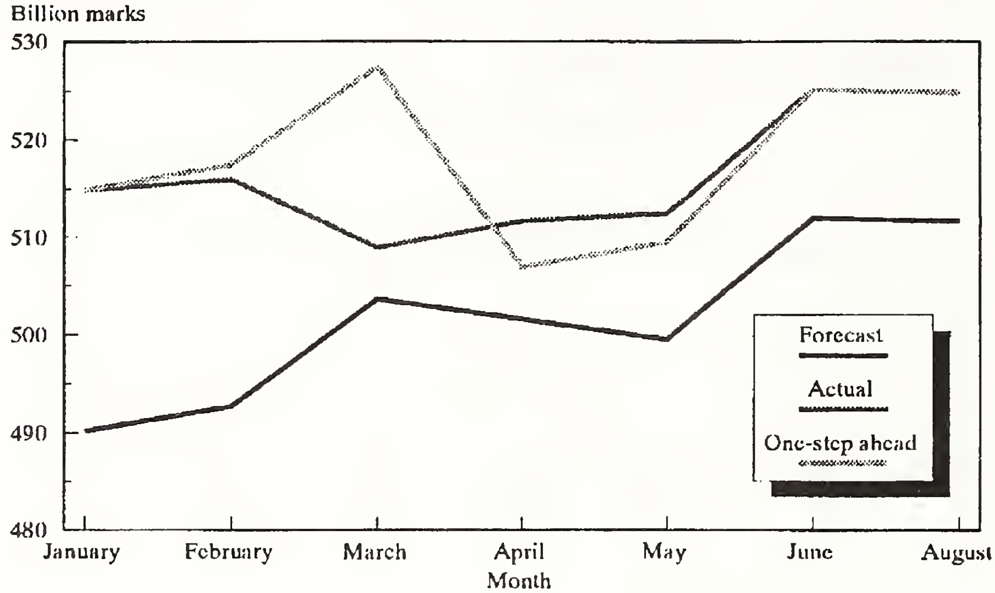


Table 1. System weights for each of the independent variables, between the input and hidden units

Variable	Total	Lag 1	Lag 3	Lag 6	Lag 12
Eurocurrency assets	100.0	14.4	27.3	26.3	32.1
Germany money market interest rates	52.5	24.2	18.0	10.9	0.5
Germany M1	52.3	19.3	8.7	3.7	20.6
Japan M1	51.6	3.1	12.4	10.7	25.3
Japan money market interest rates	47.2	19.1	15.7	5.3	7.1
U.S. CPI	34.6	9.5	17.2	2.3	5.7
U.S. money market interest rates	29.9	19.9	4.4	6.0	0.3
Mark	20.5	9.0	3.8	2.6	5.0
Germany CPI	17.1	7.6	4.0	3.8	9.4
U.S. M1	14.0	4.6	3.4	0.1	5.8
Japan CPI	4.6	1.9	3.3	0.6	1.3
Yen	4.5	0.5	5.3	0.2	1.1
U.S. Eurodollar rates	0.7	8.5	0.5	1.3	6.1

Evaluating The Age Distribution In State Population Projections

Paul R. Campbell, Population Projections Branch, U.S. Bureau of the Census

A crucial method of evaluating a forecast is to compare it with reality. From a demographic perspective, official census results may serve as a proxy measure of reality.[1] This paper compares the State population projections prepared by the Census Bureau with results from the 1990 Census. Evaluating the age, sex, and race distribution allows for the identification of areas that may need improvements or modifications in projection procedures.

This study is limited to identifying dissimilarities between the overall age structure in the projections and census results. No attempt is made to evaluate methodological sources of error affecting this comparison. For instance, the starting points of the projections (1988 population estimates) are grounded in the 1980 census results. Consequently, coverage differences in the 1980 and 1990 census results complicate the comparisons of the projections with the 1990 census. Besides enumeration errors, any comparison may be further complicated by the quality of administrative records, as well as variation in procedures used to up-date the 1980 census to the 1988 starting point .[2]

Methods.

Data - The State population projections for Whites, Blacks, and Other races evaluated in this paper are in report P-25 No. 1053.[3] The evaluation is limited to Series A of the State projections. These projections were developed using an annual cohort-component model with 1988 as the base year. Series A assumes a continuation of a modified linear trend in the annual State-to-State migration data covering 1975 to 1988.[4] First, the April 1, 1990 State population projections by sex and race in 5 year age groups were obtained by linearly interpolated between the corresponding July 1, 1989 and July 1, 1990 State population projections. Next, the April 1, 1990 State population projections were evaluated against the 1990 Census (using unpublished modified age and race results).[5] The race data examined were for Whites, Blacks, and Other races.

Evaluation Technique - The Index of Dissimilarity (D) was used to evaluate the State population projections by age, sex, and race. The D value provides a summary measure of the difference between the 1990 projected population and the 1990 Census age distribution for State, by sex and race. The percent distributions used to calculate the D value were for 5 year age groups i.e., ages 0 to 4, 5 to 9, ... 80 to 84, and 85 and over, for males and females. In this study, the differences between the percents for the corresponding age groups for the projections and Census are calculated, they are summed without regard to sign, and one-half of the sum is taken.[6] The D value is expressed as a percentage and therefore can vary from zero to 100. In other words, the projected age distributions with a D value close to zero are not very dissimilar from the Census age distribution for States. States with D values close to 100 are extremely dissimilar.[7]

Findings

Blacks and Other races most dissimilar - White males and females have the lowest D values among the 3 racial groups, indicating that projections and census age-sex distributions are very similar. The mean D values for Whites in the States were 1.9 for males and 1.7 for females, see table 1. The D values for Whites in Figures 1 and 2 varies from 0.7 to 7.4 percent. Most States with the highest D values for Whites were concentrated in the West. The 1990 Census and projected age distribution were most comparable (lowest D values) for the North East Central and Middle Atlantic States.

Table 1 Index of Dissimilarity of State Age Distributions (mean D value for States age distribution, by race and sex)

	Male	Female
Whites	1.9	1.7
Blacks	5.3	5.7
Other races	7.6	7.1

Source: Based on the State populations projections Series A for April 1, 1990 and the 1990 Census, see Appendix table 1.

Average D values for Black males and females were more than twice as high as their White counterparts. The mean D value for Blacks in the States were 5.3 for males and 5.7 percent for females. The D values for Blacks shown in Figures 3 and 4 varies from 1.1 to 37.1 percent. States with the highest D values (10 percent or more) among Black males or females have very few Blacks (less than 8,000 persons based on the 1990 census). These states were Idaho, Maine, Nebraska New Hampshire, North Dakota, South Dakota, Vermont, and Wyoming. States in the South where the Black population is concentrated tend to have the lowest D values.

The D values for Other races males and females were the highest. The mean D value for Other races in the States were 7.6 for males and 7.1 for females. The D values shown in Figures 5 and 6 varies from 1.8 to 18.7 percent. Most of the States with the highest D values have few Other races based on the 1990 Census.

Comparable level of error in State totals - Finally, the percent differences were obtained from the total projected State populations for April 1, 1990 and the census, by sex and race. By summing the percent differences regardless of the direction of the sign (plus or minus) one derives the mean absolute errors for States. The mean absolute errors for the States total population, by sex and race are close in magnitude to the mean D values for the States age distribution.[8]

Summary and Conclusions. The findings suggest that age distributions in the State population projections were not markedly dissimilar for Whites from the 1990 census (mean D values for the State equal less than 2.0 percent for either sex). In comparison, the results were twice as high for Blacks and 3 times higher for Other races.

As expected race groups which are identified as more difficult to measure accurately in the decennial censuses reflect a higher degree of dissimilarity. Nevertheless, discordance between the forecast and census age distribution may be due more or less to the failure to gauge accurately any one of the State population projection components, i.e., fertility, mortality, and migration. When information for these components are inadequate or incomplete any assumptions about future forecasts of these components will be inaccurate.

Additionally, the low national population projections and the methodological approach used to develop the base State populations all contribute to the disparities between the projections and the census age distributions.[9] These findings suggest that as the race/ethnicity categories expand in the State population projections, more attention needs to be focused on methodology and

techniques that will affect subgroups at small or low levels of geography. [10] More definitive conclusions should be available once we complete a more detailed study of the forecast errors by age.

Footnotes

1 Alternative proxy measures are sample surveys and estimates.

2 The quality of administrative records at the State level may be affected by underreporting or misclassification of registered births and deaths; school and Medicare enrollment, and Federal income tax data. These data are used to obtain subnational fertility, mortality, and migration components necessary to update the 1980 census population to 1988.

3 U.S. Bureau of the Census. 1990, ***Projections of the Population of States, by Age, Sex, and Race: 1989 to 2010.*** Current Population Reports, Series P-25, No. 1053.

4 In an earlier analysis, Series A appears to be tracking better than the other projections series when comparing state total population projections with corresponding estimates, see Paul Campbell, 1990, "Evaluation of Recent State Population Projections", in ***Federal Forecasters Conference 1990: Proceedings*** (AGES 9109).

5 The race statistics were modified to be consistent with the classification used in data sets other than the census, while the age data were adjusted to correspond with the April 1, 1990 census date. For a detailed discussion of these modifications, see Bureau of the Census, 1991, "Age, Sex, Race, and Hispanic Origin Information From The 1990 Census: A Comparison of Census Results With Results Where Age And Race Have Been Modified", 1990 CPH-L-74.

6 For a more detailed discussion and an illustration of the calculation of the index of dissimilarity, see U.S. Bureau of the Census, 1971, **The Methods and Materials of Demography**, by Henry S. Shryock, et al., U.S. Government Printing Office, p. 232.

7 Here, the index may be interpreted as the amount of change necessary for the projections to attain the same age distribution as the Census.

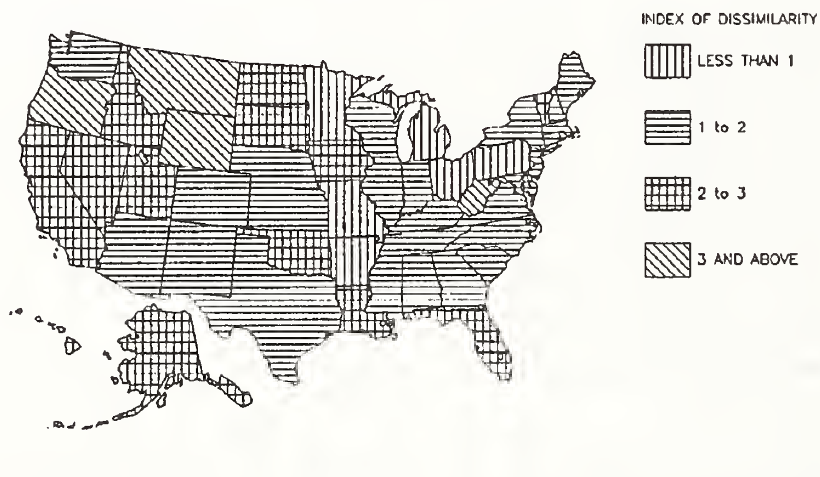
8 A non-parametric test of association, the gamma statistic was used to test for a relationship between the level of error in the total populations and the level of error in the age distribution, by sex and race for States. No relationship was found among Whites and Other races by sex (values for gamma = less than 0.2 for all groups). A very weak relationship was found among Blacks (gamma = 0.3 for males and 0.4 for females).

9 For details on the State estimates and national projections see the following reports: U.S. Bureau of the Census. 1990, ***State Population and Household Estimates: July 1, 1989.*** Current Population Reports, Series P-25, No. 1058; and 1989, ***Projections of the Population of the United States, by Age, Sex, and Race: 1988 to 2080.*** Current Population Reports, Series P-25, No. 1018.

10 The Census Bureau hopes to provide future State population projections for Whites, Blacks, and American Indians, Eskimos, or Aleuts, and Asians or Pacific Islanders, by Hispanic origin.

FIGURE 1.

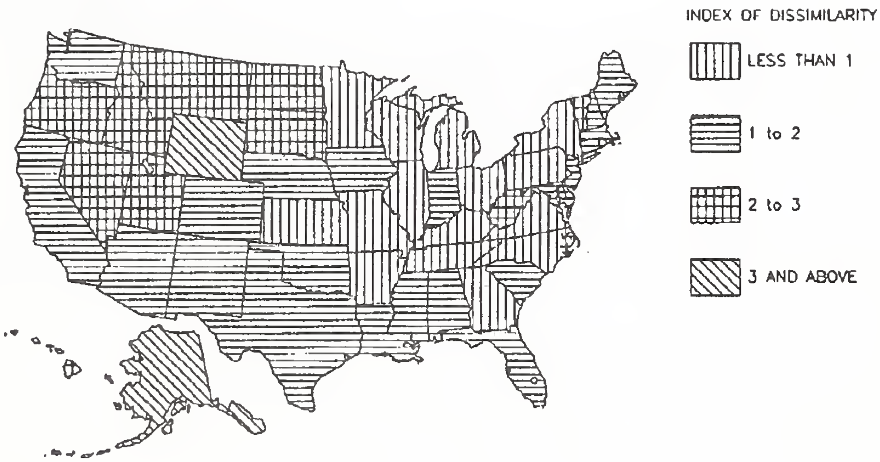
COMPARISON OF AGE DISTRIBUTIONS FOR WHITE MALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: SERIES P-25, NO. 1053.

FIGURE 2.

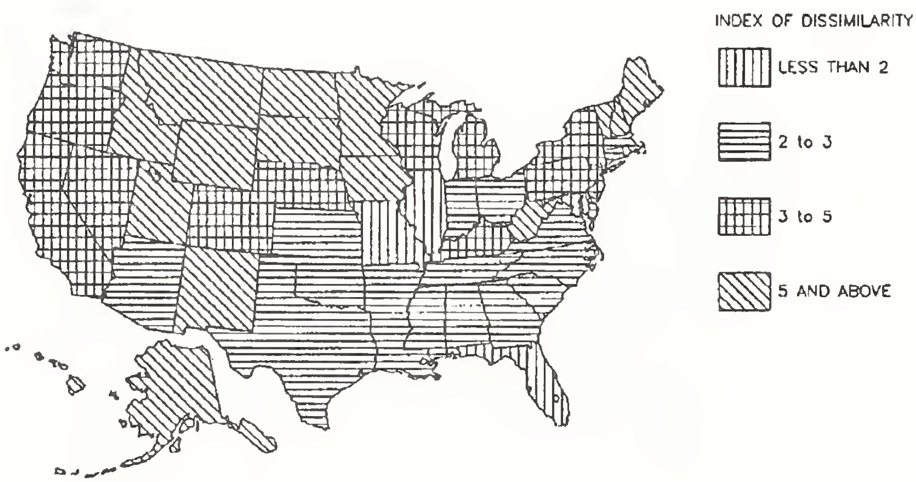
COMPARISON OF AGE DISTRIBUTIONS FOR WHITE FEMALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: SERIES P-25, NO. 1053.

FIGURE 3.

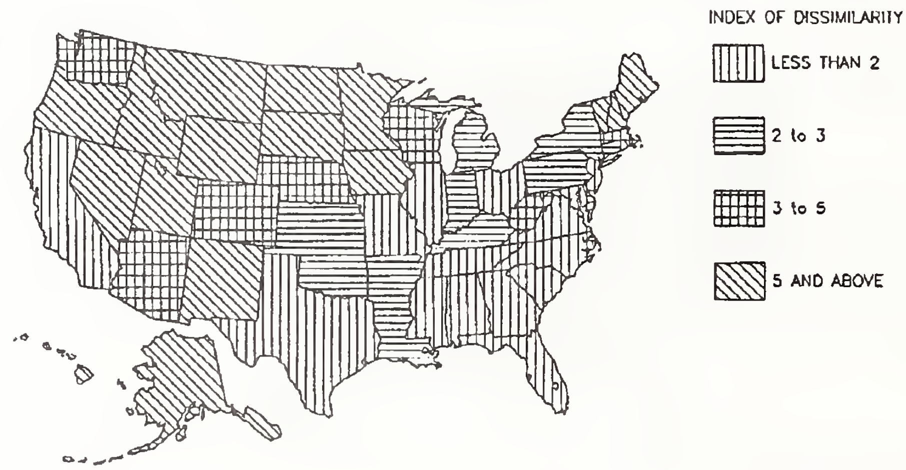
COMPARISON OF AGE DISTRIBUTIONS FOR BLACK MALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: SERIES P-25, NO. 1053.

FIGURE 4.

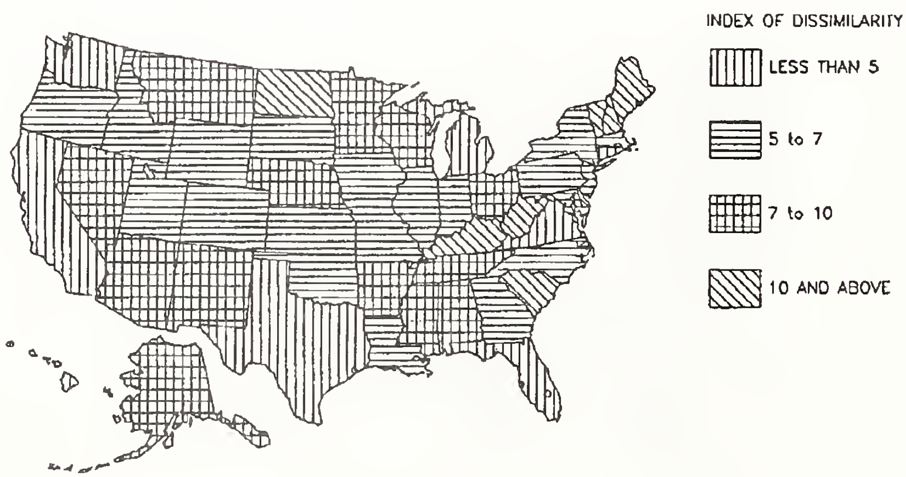
COMPARISON OF AGE DISTRIBUTIONS FOR BLACK FEMALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: P-25, NO. 1053.

FIGURE 5.

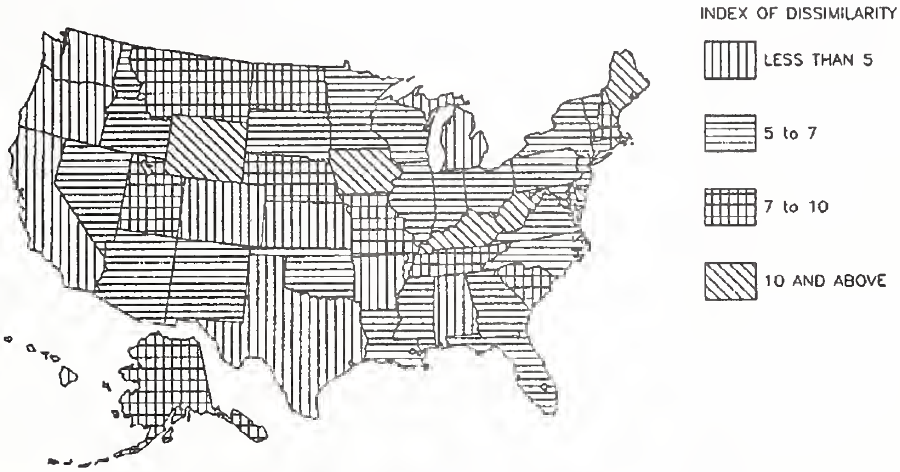
COMPARISON OF AGE DISTRIBUTIONS FOR OTHER MALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: SERIES P-25, NO. 1053.

FIGURE 6.

COMPARISON OF AGE DISTRIBUTIONS FOR OTHER FEMALES
STATE PROJECTIONS COMPARED TO 1990 CENSUS



SOURCE: SERIES P-25, NO.1053.

APPENDIX

Table 1 Index of Dissimilarity for April 1 1990 Census Population and Projected Population Age Distributions for States, by Race and Sex

State	Total		White		Black		Other races	
	male	female	male	female	male	female	male	female
Alabama	1.75	1.46	1.80	1.56	2.52	1.57	7.14	4.91
Alaska	2.04	3.34	2.24	3.43	6.37	8.38	7.89	7.18
Arizona	1.50	0.90	1.95	1.42	2.42	3.01	7.15	6.47
Arkansas	0.69	0.87	0.81	0.92	2.51	2.18	8.37	4.80
California	2.79	1.74	2.84	1.77	3.98	1.61	2.75	3.53
Colorado	1.72	1.22	1.70	1.17	3.40	3.32	5.25	4.37
Connecticut	1.97	1.53	1.77	1.42	2.74	2.30	8.65	8.05
Delaware	1.59	1.59	1.70	1.51	3.26	3.74	13.82	10.09
Dist. of Columbia	3.94	3.87	7.32	7.19	3.65	2.95	12.02	15.84
Florida	1.94	1.69	2.19	1.85	1.43	1.10	4.73	8.37
Georgia	1.29	1.11	1.08	0.84	2.24	1.85	6.86	8.19
Hawaii	2.01	1.88	2.79	2.92	7.28	7.14	1.82	2.70
Idaho	2.72	2.44	2.71	2.47	18.07	17.05	6.05	5.01
Illinois	0.98	0.67	1.15	0.74	1.79	1.10	6.04	5.65
Indiana	1.31	1.09	1.31	1.07	2.64	2.11	6.25	6.76
Iowa	2.38	1.83	2.35	1.86	5.10	5.21	6.20	11.93
Kansas	1.28	1.00	1.43	0.99	2.84	2.86	5.25	3.85
Kentucky	1.23	0.90	1.24	0.88	3.11	2.28	11.64	12.11
Louisiana	1.94	1.56	2.12	1.80	2.13	2.03	6.57	5.71
Maine	1.27	1.53	1.27	1.59	13.71	13.82	10.37	12.25
Maryland	1.26	1.38	1.16	1.40	1.84	1.54	5.43	6.60
Massachusetts	1.58	1.44	1.34	1.25	2.88	3.79	6.73	8.48
Michigan	1.18	0.96	1.00	0.81	3.33	2.28	3.87	4.25
Minnesota	1.08	0.91	0.80	0.75	5.38	5.49	8.31	6.11
Mississippi	1.51	1.27	1.09	1.40	2.81	1.74	8.70	8.83
Missouri	0.75	0.78	0.75	0.85	1.93	1.39	6.90	7.86
Montana	2.99	1.87	3.02	2.16	19.87	25.83	9.88	9.30
Nebraska	1.84	1.41	1.73	1.40	4.97	3.51	8.06	8.77
Nevada	2.57	2.80	2.65	2.59	3.52	5.91	7.02	6.05
New Hampshire	1.08	1.61	1.22	1.55	11.58	8.65	11.21	9.99
New Jersey	1.82	1.20	1.61	1.17	3.01	1.34	5.91	6.17
New Mexico	0.99	0.95	1.85	1.55	8.61	7.22	7.32	6.71
New York	2.02	1.27	1.41	0.90	3.44	2.25	6.78	6.94
North Carolina	1.20	1.01	1.04	0.78	2.36	1.72	5.51	5.50
North Dakota	2.39	1.86	2.69	2.20	14.26	27.25	15.26	8.56
Ohio	0.98	0.65	0.85	0.59	2.87	1.63	8.11	5.35
Oklahoma	1.70	1.44	2.10	1.97	2.57	2.85	5.56	5.69
Oregon	3.04	2.69	3.17	2.84	3.77	7.52	6.91	3.27
Pennsylvania	1.13	0.80	0.98	0.78	3.23	2.14	6.41	6.26
Rhode Island	1.93	1.71	1.67	1.69	6.01	3.68	8.12	8.29
South Carolina	1.48	1.25	1.30	1.18	2.59	1.74	12.09	8.71
South Dakota	2.03	1.70	2.30	2.11	7.88	37.11	8.88	5.41
Tennessee	1.33	0.80	1.23	0.79	2.74	1.88	7.27	7.82
Texas	1.04	0.95	1.16	1.10	2.60	1.54	3.87	3.21
Utah	2.47	2.30	2.70	2.40	13.07	8.35	6.37	7.80
Vermont	2.15	2.38	2.25	2.41	19.09	11.21	18.68	5.19
Virginia	1.10	0.91	1.01	0.77	2.40	1.59	4.33	6.20
Washington	1.94	1.31	1.95	1.47	4.47	4.29	4.57	3.76
West Virginia	3.04	2.11	3.12	2.22	5.28	4.65	13.42	15.46
Wisconsin	1.13	0.65	1.22	0.72	4.05	3.56	7.07	5.92
Wyoming	7.13	5.18	7.43	5.43	8.83	10.30	6.65	12.02
MEAN	1.85	1.56	1.95	1.70	5.34	5.69	7.59	7.06

Note: See text for details on data sources and methods.

The Challenge of Finding the Present — Starting Points for Forecasting the Future

Jennifer C. Day, Population Projections Branch, U.S. Bureau of the Census

Most evaluations of population projections center on criticism of errors in the results or reasonableness of the assumptions. Usually overlooked are the base or initial population characteristics and rates used to form these projections. Yet, it is difficult to project accurately if the present is inaccurately measured. These points of departure may indeed be the most challenging aspects of developing population projections.

The forthcoming set of projections created by the Census Bureau will cover projections of the population by single years of age up to 100, and highlight these new features: a 1990 census base, more race disaggregation, (specifically, white, black, Asian or Pacific Islander, and American Indian, Eskimo or Aleut), and the integration of Hispanic origin projections with the race projections. The difficulty in obtaining plausible rates of change for the initial population is compounded by these further disaggregations. Focusing on mortality rates as an example, this paper highlights evidence of this challenge.

The most reliable and current available data on mortality is gathered by the National Center for Health Statistics, published monthly in a provisional 10 percent sample of deaths by age, sex and race. Combining these deaths with the Census Bureau's most accurate population estimate for the same time period easily yields death rates. However, the death data only covers ten year age groups up to age 85 and the races of white, black, and other. Therefore, the challenge is to further segment these deaths into the necessary specific categories of age, race, and Hispanic origin.

Stratifying Age Data. Two major modifications to the age data involve converting ten year age groups to five year age groups, and distributing the 85 and over group into five year age groups up to 110. The first problem is easily resolved using another set of NCHS data organized into five year age groups. The pattern of deaths in this second data set, though two years older, is used to distribute death rates for the sample year by proportionally splitting ten year age groups into five year age groups.

The second task, however, is a bit more complicated. NCHS publishes a table of deaths for single years of age between 85 and 125. Using the pattern of death rates based on the average of the last three years of this data, the age specific death rate for the sample data over age 85 is distributed into five year age groups up to 110+.

As illustrated in Figure 1, this new pattern of advanced age death rates causes some skepticism concerning the reliability of data of the extremely aged. The downturn in the oldest age groups' death rates implies an decreasing probability of death as age increases.

Several factors may contribute to this problem of data accuracy and reliability. Age data, especially in the very old, may have some serious misreporting. The national registration system of births was not created until early in the twentieth century and was not consistently used in all states until 1933. Spurious dates of birth and ages may be expected for those individuals born prior to this. Moreover, exaggeration of age, especially in the oldest ages, is not uncommon. In addition, death statistics are subject to deficiencies due to inaccuracies, incompleteness, and timeliness in reporting. More specifically, underregistration of deaths by age, misreporting of age, and not reporting are problematic points for accurately determining death rates by age. In fact, unreporting of deaths combined with data entry or clerical error distorts the data for these oldest ages more so than data for younger ages where the population is not so small. Therefore, these suspect death rates are probably due to a substantial overstatement of age combined with an understatement of mortality.

Modification of this downward trend of mortality requires use of alternative death rates provided by the Social Security Administration based on edited Medicare data. Applying this pattern of death rates for 5 year age groups over 85, yet still maintaining the same total number of deaths for ages over 85, the new death rates for the oldest ages reflect trends which appear more plausible. See Figure 2.

Disaggregating Race Data. For years, race data categorized by White, Black, and the residual "other" was sufficient. However, we now require the "other" group to be disagreggated into two specific races: Asians or Pacific Islanders and American Indian, Eskimo, or Aleuts.

Splitting the "other" race group requires sufficient data about at least one of the two groups. Although NCHS does not publish detailed death data for either of these two race groups by age and sex, the Indian Health Service does receive this detailed information from NCHS.

By using the American Indian death data from the Indian Health Service, the Other races' deaths can be allocated between the two groups. The most current Indian data available, however, was from 1988. Therefore, the American Indian deaths were subtracted from the "Other" deaths for the year 1988, the remainder being assumed Asians. Then, this proportion of "other" deaths for American Indians and Asians in 1988 was used to distribute the 1990 total "Other" deaths by age, sex, and race.

Similar to the problems with the specific "other" race data, data for Hispanics are not completely available. The most recent death data, from 1988, are collected by NCHS from 26 states—these states representing about 82 percent of the Hispanic population. To estimate the total number of Hispanic deaths, the number of deaths was increased for each age and sex group by dividing each subgroup by .82. Then, by using the white change in death rates between 1988 and 1990, the Hispanic death rates were forwarded to 1990 rates.

As pictured in Figures 3 and 4, the age specific death rates for API, American Indians, and Hispanics, especially in comparison to the white and black rates, are quite low. According to these rates, it appears that Asians, American Indians, and Hispanics experience better mortality than whites.

Although these rates are possible, intuitive judgement questions the appropriateness of these rates. Though the methodology used to derive these surprising results may be faulty, obviously there are quirks in the data themselves. More specifically, inherent confusion regarding race definition and identification contributes to the unreliability of segmented race data, especially when using multiple sources of data. For instance, the Census Bureau relies on self identification in its ten year count. However, on both birth and death certificates, which record a person's existence in the world, race is assigned by another individual. The complexity of this race problem gets even more involved when dealing with children of mixed races where identification may be very erratic.

Basically, we are left in a bit of a quandary — that is,

either we accept the race mortality data from NCHS and the resulting rates, or we consider the rates implausible and change the race numerators to reflect more acceptable rates. However, the accuracy of projections are judged by comparisons with future data from NCHS on births and deaths. Therefore, if we don't accept these base data as given, we unwittingly chance appearing wrong by these standards.

Summary. As demonstrated here, forecasting the population involves confronting and resolving initial data problems in order to develop a solid foundation. This task of accurately portraying the current population and its characteristics demands evaluation at each step, considering multiple sources of data for appropriateness and distortions. Finding good data, that is current, accurate, consistent, and reliable, though not as glamorous as creating assumptions, is the first requirement for developing good projections.

Figure 1

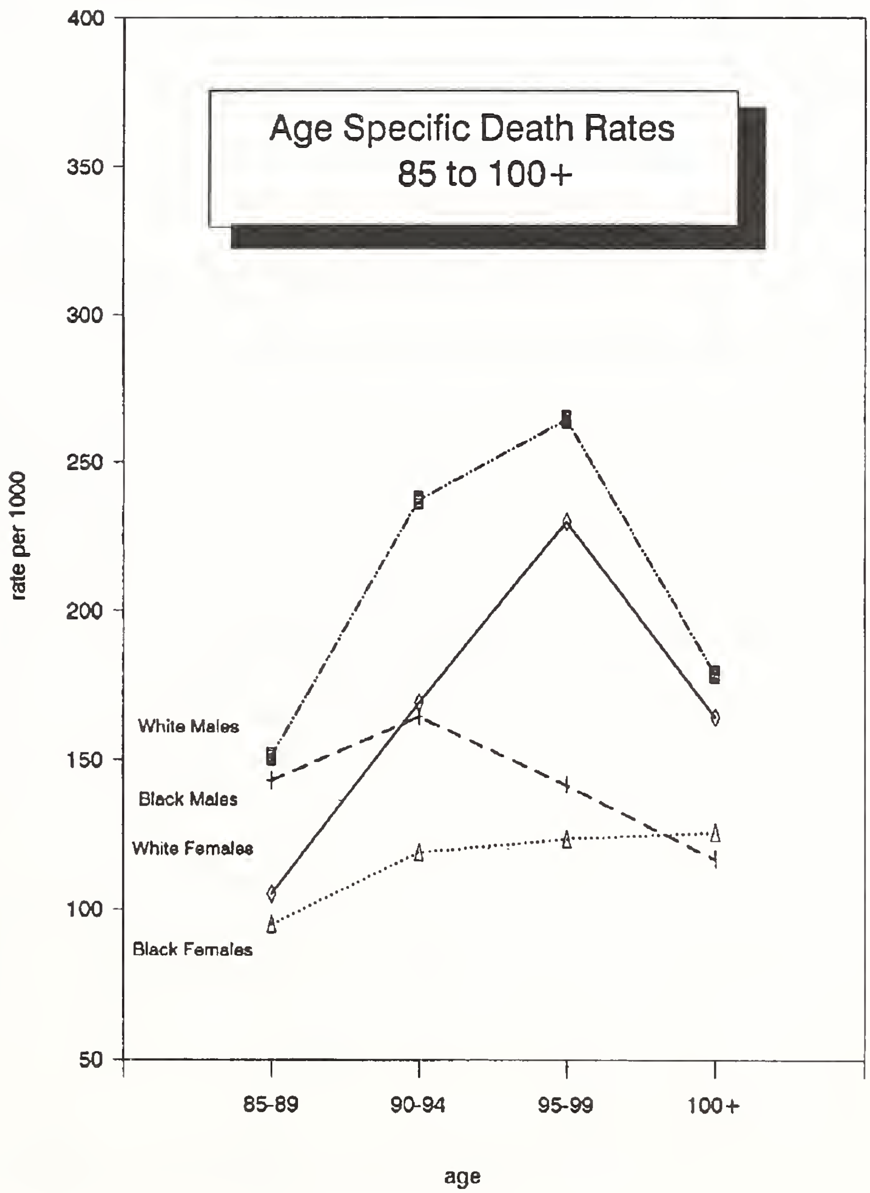


Figure 2

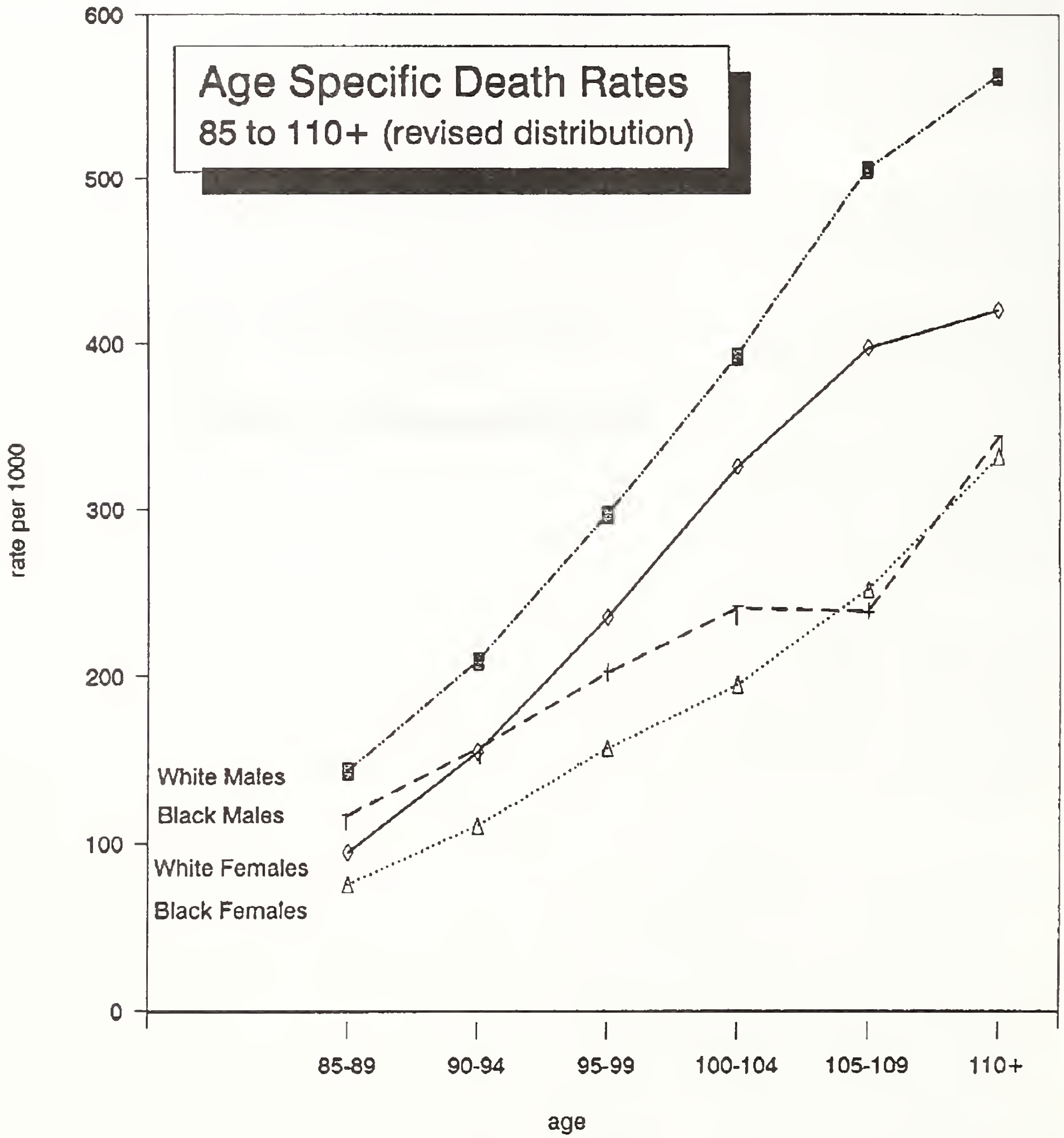

Age Specific Death Rates
85 to 110+ (revised distribution)
600
500
400
300
200
100
0
rate per 1000
White Males
Black Males
White Females
Black Females
85-89
90-94
95-99
100-104
105-109
110+
age

Figure 3

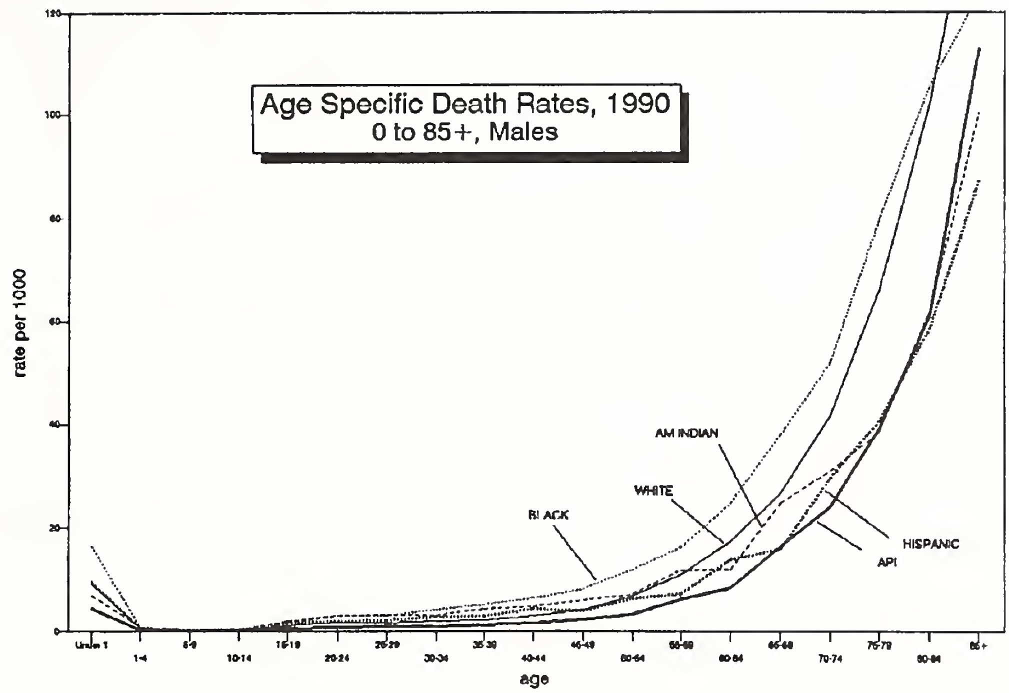
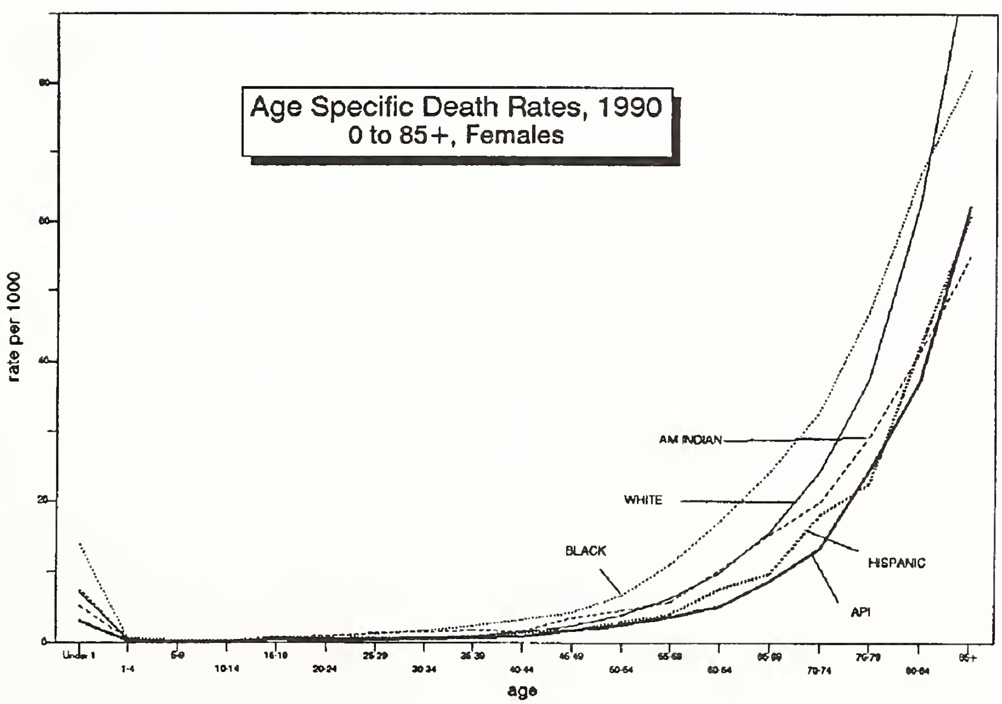


Figure 4



Early Estimates of Education Statistics: How Do We Know if They're Useful?

Bob Burton, National Center for Education Statistics, U.S. Department of Education

Summary. Typical surveys and censuses provide estimates--in either the technical or the non-technical sense of the word--and, at least in the case of sample surveys, confidence intervals around these estimates. However, there is no way to measure the accuracy of any particular estimate, since the true values are never known.

Early estimates, as this phrase is generally used, are different in an important way: They can be subsequently checked against actual values when these values become available. (Note that it is better to speak of "actual" values than of "true" values in this context, since the later data may themselves be inaccurate, i.e., may not be equal to the population values.)

Because of this difference, it is an easy matter to evaluate the accuracy of early estimates. However, accuracy, in the sense of predicting levels, does not seem to be the major function of early estimates. They serve, rather, as an early indicator of change over time, both in the narrow sense of change from the immediately preceding value in a time series, and in the broader sense of a pattern over a multi-year period.

In this situation, it is better to evaluate early estimates in terms of their usefulness to the intended audience, rather than in terms of accuracy per se. The objectives of this report are to:

1) Develop qualitative criteria for assessing the usefulness of early estimates in the context of an ongoing time series;
2) Quantify these criteria, so as to yield well-defined procedures and rules;
3) Illustrate the use of the procedures and rules with actual data.

The data chosen for illustrative purposes are the actual and early estimated total counts of students, teachers, and graduates from the Common Core of Data (CCD). Although the main purpose is to demonstrate the use of the proposed procedures, it is gratifying to note that the early estimates do turn out to be "useful" in almost all cases.

1. USEFULNESS OF EARLY ESTIMATES

THE OBJECTIVE OF THIS SECTION IS TO DEVELOP GROUND RULES FOR DECIDING WHETHER AN EARLY ESTIMATE--OF A QUANTITY FOR WHICH A TIME SERIES HAS BEEN MAINTAINED FOR SEVERAL YEARS--IS OR IS NOT USEFUL. IN SUMMARY, THE ARGUMENT TO BE PRESENTED HERE IS THAT ACCURACY OF THE ESTIMATE ITSELF IS NOT A SUFFICIENT CRITERION FOR EVALUATING AN EARLY ESTIMATE. IT IS NECESSARY TO EXAMINE THE ACCURACY OF THE IMPLIED CHANGE IN THE QUANTITY BEING MEASURED, BOTH AS A SIMPLE CHANGE FROM ONE YEAR TO THE NEXT AND AS A COMPONENT OF A LONGER-TERM TREND.

1.1. THE RELATIVE ERROR OF AN EARLY ESTIMATE

Consider a time series that has been maintained for several years, and for which there is an unavoidable delay in obtaining the actual new value for each successive year. Suppose, however, that it is possible to develop an "early estimate" each year, and that there is reason to believe that the early estimated values are "valid", or "reliable", or "accurate", or whatever word one wishes to use to convey the message that each early estimate is expected to be close to the corresponding actual value.

At some time after the early estimate is developed (and, presumably, published and disseminated), the actual value becomes available. Of course, it is then possible to see just how close the early estimate was to the actual value.

Suppose, to make this concrete, that the numbers in question, for Year t, are as follows:

	Actual Value	Early Estimate	Relative Error
Year t	40.0	40.20	0.50%

In the absence of any further relevant information, most people would characterize a relative error of one-half of one per cent as "small", and would say that the early estimate is close enough to the actual value to be itself characterized as a "good" estimate. This would seem to put the matter to rest.

1.2. USEFULNESS IN THE CONTEXT OF SHORT-TERM CHANGE

The general reason to maintain time series is to monitor trends: patterns of change over a period of time. For an annual series, the most basic pattern is that observed over a two-year period.

In this context, an early estimate is useful if it leads to an estimated change that is itself "close" to the actual change. Suppose, for example, that in Year t-I the actual value was 39.8. It would then be appropriate to look at the following table:

	Actual Value	Early Estimate	Relative Error
Year t-1	39.8		
Year t	40.0	40.2	0.50%
Change	0.50%	1.01%	100%

The early estimate indicates an increase of 40.2 minus 39.8, or 0.4, or, in relative terms, about 1.01%. The actual increase turns out to be only 0.2, or about 0.50%. The derived estimate of change is exactly twice as large as the actual change; In other words, the relative error, in terms of change, is 100%.

It would appear that the matter has not been laid to rest, since most people would characterize a relative error of 100% as "large", and would therefore characterize the estimate itself as "bad".

Suppose that the previous year's value was 39.0, rather that 39.8. Then the table would appear as follows:

	Actual Early Value	Relative Estimate	Error
Year t-1	39.0		
Year t	40.0	40.2	0.50%
Change	2.56%	3.08%	20%

How would we characterize the estimate in this case? Probably, since we recognize that it is more difficult to estimate change than to estimate level, we would tend to consider this estimate "good".

Our immediate concern, however, is not with attaching evaluative adjectives to estimates. The point to be made is that it is not wise to evaluate an early estimate in terms of how well it predicts level; It Is necessary to see how well it predicts change,

since this is the context in which the audience will be viewing the early estimate. Another way to think of this is that an estimate may be good in the traditional sense of being close to the true value, but still not be useful to the audience at which it is aimed.

1.3. USEFULNESS IN THE CONTEXT OF LONGER-TERM TRENDS

We can expect that the audience for our early estimates will be looking at more than just the immediately preceding actual value. Suppose, to extend the first example in the preceding section, that the following data are available:

	Actual Value	Early Estimate	Relative Error
Year t-5	36.1		
Year t-4	37.2		
Year t-3	37.9		
Year t-2	38.7		
Year t-1	39.8		
Year t	40.0	40.2	0.50%
Change	0.50%	1.01%	100%

With this additional information, we can note that the annual changes from Year t-5 to Year t-1 are 1.1, 0.7, 0.8, and 1.1. The average of these four annual changes is about 0.9, and we would probably expect the change from Year t-1 to Year t to be near this value.

In this context, the early estimate is a "surprise": It indicates that the rate of increase is expected to drop to about half of the value that had been observed over the past several years.

In fact, the increase from Year t-1 to Year t is 0.2, which is even smaller than that implied by the early estimate. In this situation, we would certainly want to characterize the early estimate as "good", since it correctly provides a clear signal that the rate of increase is slowing down.

1.4. A GENERAL DEFINITION OF "USEFULNESS"

The following ground rules are proposed for determining the usefulness of an early estimate:

A. The relative accuracy of the estimate is not, in and of itself, a determinant of usefulness;

B. An estimate that accurately predicts change from the preceding year is useful;

C. An estimate that accurately classifies change from the preceding year, in the context of previously observed changes, is useful;

D. An estimate that neither predicts change accurately nor classifies this change accurately is not useful.

With these ground rules, both of the estimates that we have used as examples are useful ones. One is useful, assuming that a 20% relative error is deemed sufficiently accurate, because it satisfies Condition B. The other is useful, even though it is off by 100% in terms of predicting the magnitude of the change, because it satisfies Condition C: It accurately leads to the conclusion that change will be smaller than would be expected on the basis of previous years' changes.

2. PROPOSED CRITERIA FOR EVALUATING AN EARLY ESTIMATE

THE GROUND RULES DEVELOPED IN THE PRECEDING SECTION REQUIRE DEFINITION AND QUANTIFICATION IF THEY ARE TO SERVE AS TOOLS FOR EVALUATING EARLY ESTIMATES. THE OBJECTIVE OF THIS SECTION IS TO SPECIFY EXACT PROCEDURES FOR DECIDING WHETHER AN EARLY ESTIMATE PROVIDES AN "ACCURATE PREDICTION OF CHANGE" AND/OR "AN ACCURATE CLASSIFICATION OF CHANGE". THE PROPOSED PROCEDURES, ALTHOUGH EXACT, ARE ALSO ARBITRARY, AND SHOULD BE VIEWED AS SUBJECT TO CHANGE. THIS SECTION ALSO ADDRESSES THE DISTINCTION BETWEEN EVALUATING AN ESTIMATE AND EVALUATING AN ESTIMATOR, AND DISCUSSES THE DIFFICULTY IN DOING THE LATTER IN THE CASE OF EARLY ESTIMATES.

2.1 PROPOSED CRITERION FOR AN "ACCURATE PREDICTION OF CHANGE"

We have now defined "usefulness" in terms of an accurate prediction of change and/or an accurate classification of change in the context of past changes. The next step is to quantify these concepts.

The accuracy of a predicted change is a relatively easy concept to quantify, being analogous to the concept of the accuracy of a predicted level. However, as mentioned in the preceding section, it is always more difficult to predict change accurately, since a small error in one or the other of two estimated values can lead to a very large error in the estimated difference between the two. As a result, it does not seem reasonable to insist on a relative error of no more than 2% or 5% in order to qualify an estimated change as "accurate".

We will tentatively propose a relative error of 30% as the cutoff between "accurate" and "inaccurate". This--or any other--cutoff is obviously arbitrary.

It should be noted that there is no compelling reason to set any cutoff value. In evaluating the adequacy of a estimated change, we could, as is typically done with an estimated level, not commit ourselves to an a priori criterion, but settle for computing (and publishing, along with the primary estimate) a standard error or a confidence interval. However, it seems worthwhile, since we don't usually think in terms of acceptable errors for estimated changes, to at least propose a tentative criterion to anchor our thinking.

A second point to note is the obvious one that, as true change approaches zero, the relative error of an estimated change, given any nontrivial error mechanism, approaches infinity. Suppose, to make this concrete, that the true values for years t-1 and t are 39.000 and 39.001, respectively, and that the early estimated value for year t is 39.010. The error in the estimated change is then 0.009, and the associated relative error is 900%. This is not an acceptable error.

Our intention is that this estimate will, under normal circumstances, pass the second criterion: that of accurately predicting the classification of change relative to past changes. What we mean by this is that there have been other years in the recent past in which the true change has been distinctly larger, in absolute value, than 0.001, so that we would classify both the latest estimated change and the latest true change as "no appreciable change", and therefore consider the estimate to be accurate in this

regard.

(If the changes from year to year are always of the order of a few thousandths of a unit, then the estimate is indeed "inaccurate", i.e., not of any use to the target audience.)

2.2 PROPOSED CRITERION FOR AN "ACCURATE CLASSIFICATION OF CHANGE"

By "classification of change", we refer to the process of taking a change from one year to the next, and assigning one of several possible labels to it. A particularly easy way to do this, for example, would be to characterize a change as positive, negative, or no change at all. Then, if an early estimate were greater than the previous year's value, and if the actual value were also greater than last year's value, we would consider the estimate to be accurate: It implied an increase, and an increase actually occurred.

Our intention, however, is not merely to classify change, but to do so "in the context of past changes". To meet this requirement, the following method is proposed:

1) Take the changes for the last eight pairs of years for which true values are available, and arrange them in ascending order;
2) Set two cutoff values, the first equal to the third change on the ordered list, and the second equal to the sixth change;
3) Classify the current estimated change as "low", "moderate", or "high", depending on whether it is below the first cutoff, at or above the first but below the second, or at or above the second, respectively;
4) Classify the current true change in the same way; and
5) Characterize the estimate itself as accurate if the two classifications are the same, and as inaccurate if they are not the same.

This criterion is clearly arbitrary in that there are many ways to classify change and to thereby decide whether an implied change is useful in the context of past points in a time series. This is simply one procedure for doing so. As with the first criterion, it provides an anchor point, and might well be changed if it turns out to be too strict, or to be too lenient, or to miss some salient aspect of trends over time.

2.3 THE PROBLEM OF EVALUATING THE ESTIMATOR

The two proposed criteria for evaluating an early estimate do just that: they provide rules for assessing the usefulness of a particular estimated value. This is in distinction to the standard treatment of sampling error, which provides rules for judging the precision of an estimator, rather than the accuracy of any one realization of that estimator.

Given one realization of an early estimate, there is no way to evaluate the estimator or, equivalently, the process that led to the estimate. If the estimate is exactly on target, to state this in another way, there is no statistical basis for expecting it to be on target again--or even close to its true value--in the following year. As a consequence, there is no statistical basis for deciding whether or not to continue producing and publishing an early estimate after its first year of use.

After an early estimator has been produced two or more times, statistical inference becomes possible. However, it is certainly risky to perform such inference based on a handful of data points.

We choose not to give rules in this situation. If an early estimation process has been in use for, say, three years, and the resulting estimates have been deemed "useful" in two of these years and "not useful" in one, it is left to others to develop appropriate statements regarding the overall usefulness of the process.

It should be noted here that aggregated early estimates, e.g., for the nation, are typically constructed by summing estimates for lower-level units, such as states or institutions. Given the disaggregated estimates and the corresponding true values, it is possible to use one year's worth of data to assess the accuracy of the process itself. However, we prefer to avoid the complications that this approach would entail, and to focus, at least in this paper, on the concepts and rules that have been developed thus far.

3. APPLICATION TO ESTIMATED PUBLIC SCHOOL MEMBERSHIP

IN THIS SECTION, THE CRITERIA DEVELOPED ABOVE ARE APPLIED TO THE EARLY ESTIMATES OF PUBLIC SCHOOL FALL MEMBERSHIP, WHICH ARE GENERATED AS PART OF THE COMMON CORE OF DATA. DATA ARE AVAILABLE FOR THREE YEARS: 1987 THROUGH 1989. THE EARLY ESTIMATES TURN OUT TO BE USEFUL FOR ALL THREE YEARS. FOR TWO OF THE YEARS, THE PREDICTED CHANGE IS ACCURATE; FOR ONE YEAR, THE PREDICTED CHANGE IS NOT ACCURATE, BUT THE CLASSIFICATION OF THE CHANGE IS ACCURATE.

3.1 BACKGROUND AND DATA VALUES

Early estimates of key statistics for public elementary and secondary education have been produced and published for four years: in December of 1987, 1988, 1989, and 1990. Key statistics include primary data elements (membership, number of teachers, number of high school graduates, revenues, and current expenditures), as well as derived data elements (pupil/teacher ratio, per pupil revenue, and per pupil expenditure).

The membership figures are totals for grades prekindergarten through 12, and are intended to reflect fall membership at the beginning of the school year. The 1987 early estimate, for example, is an estimate released in December, 1987, of the membership that was recorded throughout the nation in October, 1987.

An early estimate is developed for each state and for the District of Columbia. These 51 estimates are then added to obtain the early estimated membership for the nation. Our concern in this report is only with the national figures.

The actual membership figures are published about one year after data are recorded at the local level, e.g., for 1987 data, in the fall of 1988. Thus, at the present time (April, 1991), actual national totals are available for all years through 1989, and it is possible to evaluate the early estimated national memberships for 1987, 1988, and 1989.

The relevant data are shown in Table 1. These data include actual counts for 1978 through 1989, and early estimates for 1987 through 1989. Table I also provides the actual relative increases from 1979 through 1989, the estimated relative increases from 1987 through 1989, and the relative errors in the latter set of increases.

3.2 EVALUATION RESULTS

The early estimated membership for 1987 implied an increase, from the 1986 value, of 0.91%. The actual increase turned out to be 0.47%. Thus, the relative error of the estimated increase was (0.91 - 0.47) / 0.47, or 94%. Since this error is greater than 30%, the early estimate cannot be judged useful on the basis

of accurate prediction of change, and it is necessary to look at the second criterion: accurate classification of change.

For the eight pairs of successive years ending in 1986, the annual changes ranged from a low value of -2.19% to a high value of 0.83%. When these eight changes are ranked, from the lowest to the highest,the values in positions three and six are -1.75% and 0.11%, respectively.

Since 0.91% is greater that 0.11%, the early estimate indicated that the change from 1986 to 1987 would be relatively high. The actual change, 0.47%, while lower than the estimated change, is still above 0.11%. It is, in other words, relatively high. Therefore, the estimated change classification is accurate, and the early estimate itself should be considered useful.

The 1988 early estimated membership implies a change of 0.43% from 1986. The actual change was 0.42%, and the relative error of the estimated change was 2.4%. Since this is less than 30%, the predicted change should be considered accurate, and, again, the early estimate is a useful one. (Since the first criterion is met, there is no need to look at the second criterion.)

For 1989, the situation is the same as for 1988, in that the relative error of the estimated change was 24.6%, which is less than 30%. Therefore, the early estimate is a useful one.

4. APPLICATION TO OTHER CCD EARLY ESTIMATES

IN THIS SECTION, THE SAME CRITERIA ARE APPLIED TO TWO OTHER SETS OF CCD EARLY ESTIMATES: NUMBER OF TEACHERS AND NUMBER OF GRADUATES. THE FORMER EARLY ESTIMATES ARE JUDGED USEFUL FOR ALL THREE OF THE YEARS FOR WHICH COMPLETE DATA ARE AVAILABLE. THE LATTER ARE JUDGED USEFUL FOR TWO OF THE THREE YEARS.

4.1 EARLY ESTIMATES OF NUMBER OF TEACHERS

The relevant data for number of teachers are shown in Table 2. In 1987 and 1989, the estimated changes for this variable were quite accurate: The errors were -8.57% and 7.32%, respectively. The early estimates for these two years are therefore judged to have been useful.

In 1988, the error in the estimated change was high: -54.05%. The cutoff points for the previous eight changes are 0.05% and 1.68%. The actual change for 1988 was 1.62%, which is between the two cutoffs, and is therefore classified as moderate. The estimated change of 0.75% is also moderate, leading to the conclusion that the 1988 early estimate should be thought of as useful.

Thus, as was the case with fall enrollments, the estimated numbers of teachers were useful in each of the three years.

4.2 EARLY ESTIMATES OF NUMBER OF GRADUATES

The data for number of graduates are shown in Table 3. The estimated changes for 1987 and 1988 were quite accurate, with errors of -9.80% and -6.25%, respectively. Therefore, the early estimates for these two years are considered useful.

The estimated change for 1989 was in error by -97.87%. Furthermore, this estimated change was in the high range, whereas the true change was moderate. Therefore, the early estimate for 1989 fails to meet either criterion, and is not considered useful.

Table 1: FALL MEMBERSHIP IN PUBLIC ELEMENTARY AND SECONDARY SCHOOLS: ACTUAL AND EARLY ESTIMATED VALUES, ACTUAL AND EARLY ESTIMATED CHANGES, AND RELATIVE ERRORS OF ESTIMATED CHANGES

Year	Actual M'ship	Actual Change	Ranked Change	Est'd M'ship	Est'd Change	Relative Error
1978	42,550					
1979	41,645	-2.13%	2			
1980	40,918	-1.75%	3			
1981	40,022	-2.19%	1			
1982	39,566	-1,14%	4			
1983	39,252	-0.79%	5			
1984	39,295	0.11	6			
1985	39,509	0.54	7			
1986	39,837	0.83%	8			
1987	40,024	0.47%		40,200	0.91%	94.12%
1988	40,192	0.42%		40,196	0.43%	2.38%
1989	40,526	0.83%		40,608	1,04%	24.55%

Notes:

1) Membership values are in thousands. Actual memberships are from the Digest of Education Statistics (1990) for years through 1985, and from the Early Estimate publications for later years.

2) Ranked changes are shown only for the eight years immediately preceding 1987, since it is only for the 1987 early estimate that these changes are needed.

Table 2: NUMBER OF TEACHERS IN PUBLIC ELEMENTARY AND SECONDARY SCHOOLS: ACTUAL AND EARLY ESTIMATED VALUES, ACTUAL AND EARLY ESTIMATED CHANGES, AND RELATIVE ERRORS OF ESTIMATED CHANGES

Year	Actual M'ship	Actual Change	Ranked Change	Est'd M'ship	Est'd Change	Relative Error
1978	2,206	-				
1979	2,183	-1.04%				
1980	2,184	0.05%	3			
1981	2,185	-2.70%	1			
1982	2,125	-0,19%	2			
1983	2,121	0.24%	4			
1984	2,168	1.98%	8			
1985	2,207	1.80%	7			
1986	2,244	1.68%	6			
1987	2,279	1.56%	5	2,276	1.43%	-8.57%
1988	2,316	1.62%		2,296	0.75%	-54.05%
1989	2,357	1.77%		2,360	1.90%	7.32%

Notes:

1) Numbers of teachers are in thousands. Actual numbers are from the Digest of Education Statistics (1990) for years through 1985, and from the Early Estimate publications for later years.

2) Ranked changes are shown only for the eight years immediately preceding 1988, since it is only for the 1988 early estimate that these changes are needed.

Table 3: NUMBER OF GRADUATES FROM PUBLIC SECONDARY SCHOOLS:
ACTUAL AND EARLY ESTIMATED VALUES, ACTUAL AND EARLY ESTIMATED CHANGES, AND RELATIVE ERRORS OF ESTIMATED CHANGES (current-year estimates)

Year	Actual Grads	Actual Change	Ranked Change	Est'd Grads	Est'd Change	Relative Error
1978	2,825					
1979	2,817	-0.28%				
1980	2,748	-2.45%				
1981	2,725	-0.84%	5			
1982	2,705	-0.73%	6			
1983	2,598	-3.96%	1			
1984	2,495	-3.96%	2			
1985	2,414	-3.25%	3			
1986	2,382	-1.33%	4			
1987	2,433	2.14%	7	2,428	1.93%	-9.80%
1988	2,497	2.63%	8	2,493	2.47%	-6.25%
1989	2,450	-1.88%		2,496	-0.04%	-97.87%

Notes:

1) Numbers of graduates are in thousands. Actual numbers are from the Digest of Education Statistics (1990) for years through 1985, and from the Early Estimate publications for later years.

2) Ranked changes are shown only for the eight years immediately preceding 1989, since it is only for the 1989 early estimate that these changes are needed.

Using Principal Components in Time Series Modeling and Forecasting of Age-Specific Mortality Rates

William Bell and Brian Monsell, U. S. Bureau of the Census

1. Introduction Population projections are typically made within a cohort-component framework by projecting the basic demographic components -- births, deaths, and migration. These components, in turn, are often projected by forecasting corresponding age-specific rates. This approach typically leads to a forecasting problem of high dimension, with demographic rates for a large number of ages to forecast.

To reduce the dimensionality in forecasting age-specific fertility rates, Bozik and Bell (1987) developed a principal components approach, which we review in section 2. Taking the first J principal components defines a linear transformation of the data with reduced dimension (for J less than the dimension of the data, i.e. number of ages). Such a transformation yields a linear approximation of the full set of original time series with minimum squared error. Similar work is done by Sivamurthy (1987) and Carter and Lee (1990).

As will be seen in the analysis of central death rates in section 3, however, all 22 principal component series are forecastable to some degree, so that leaving any of them out of the approximation has some deleterious effect on the forecasts. Also, ignoring the error in a reduced dimension approximation results in forecast error variances that are too low because of this error ignored. Therefore, in thispaper we develop an approach to model and forecast the full set of principal component series. This provides a method free of approximation error for producing point and interval forecasts of the full set of time series of demographic rates. While we have no approximation error, we also have not reduced the dimensionality of the forecasting problem either. The advantage of the approach is that transforming to the principal component time series greatly simplifies the structure of the modeling and forecasting problem, making it feasible to develop a multivariate time series model for the full set of principal component series. In contrast, directly developing a multivariate model for the full set of original demographic rates would be practically impossible.

In section 2 we review the principal components approach, including a discussion of how one can develop a multivariate time series model for the full set of principal components. We also show how this model can be used to produce point and interval forecasts for the original series of demographic rates. In section 3 we apply the approach to forecasting age-specific central death rates of U.S. white females. Section 4 provides conclusions.

2. Principal Components Approach We briefly outline the principal components approach, which is discussed in more detail in Bozik and Bell (1987). Let the age-specific demographic rates we wish to forecast (fertility or mortality) for individuals of agegroup k in year t be denoted as r_{kt}. Actually, it will typically be desirable to let r_{kt} be some transformation of the original series of demographic rates. In section 3, we let $r_{kt} = \log(m_{kt})$ where the m_{kt} are the age-specific central death rates to white women in the U.S. Taking logarithms assures that, after inverting the transformation for the forecasts of the r_{kt}, the resulting demographic rate forecasts (point and interval) remain positive.

Suppose now that we have a given set of constants λ_{kj} for the span of age groups $k = 1, \cdots, K$ and for $j = 1, \cdots, J$, where J is the number of linear functions of the data we use to approximate $\underline{r}_t$, where $\underline{r}_t = [r_{1t} \cdots r_{Kt}]$. Let $\Lambda = [\lambda_{kj}]$, which has dimension K x J. The principal components approach finds coefficients $(\beta_{1t} \cdots \beta_{Jt})' = \underline{\beta}_t$ to

$$\min \sum_k [r_{kt} - (\beta_{1t}\lambda_{k1} + \cdots + \beta_{Jt}\lambda_{kJ})]^2 = \min |\underline{r}_t - \Lambda\underline{\beta}_t|^2$$

both with respect to $\underline{\beta}_t$. With the columns of $\Lambda = [\underline{\lambda}_1 \cdots \underline{\lambda}_J]$ restricted to be orthonormal, the solution is found at

$$\underline{\beta}_t^* = (\Lambda'\Lambda)^{-1}\Lambda'\underline{r}_t = \Lambda'\underline{r}_t$$

For this choice of $\underline{\beta}_t$, we then pick Λ to minimize the aggregate approximation error for any J, that is to

$$\min \sum_t |\underline{r}_t - \Lambda\underline{\beta}_t^*|^2$$

with respect to Λ. The result is that the columns of Λ are the eigenvectors of the sum of squares and cross products matrix of the data, $\Sigma_t \underline{r}_t\underline{r}_t'$. The approximation to $\underline{r}_t$ is $\underline{\beta}^*_t = \Lambda\Lambda'\underline{r}_t$. Notice if J = K then Λ is square and orthogonal, with $\Lambda'\Lambda = \Lambda\Lambda' = I$, so $\Lambda\underline{\beta}^*_t = \underline{r}_t$ and there is no approximation error. For simplicity of notation in what follows, we shall use $\underline{\beta}_t$ instead of $\underline{\beta}^*_t$ to denote the principal components regression coefficients.

2.1 Multivariate Modeling of All the Principal Components In Bozik and Bell (1987) a weighted principal components approach was used to approximate fertility rates for 33 ages using a principal components transformation of dimension 5, assuming the approximation error could be ignored in forecasting. Here we show how to use all the principal components. Then $\underline{\beta}_t$ is simply a nonsingular linear transformation of $\underline{r}_t$, and there is no approximation error, but there is no reduction in dimension either. Why would we prefer to model the K x 1 $\underline{\beta}_t$ instead of the K x 1 $\underline{r}_t$? Because $\underline{\beta}_t$ turns out to be much simpler to model than $\underline{r}_t$. Notice that, through their derivation as a solution to the approximation problem, $\beta_{1t}, \cdots, \beta_{Kt}$ are ordered in importance with regard to their contributions to variation in the original demographic rates r_{kt}. Thus, most of the attention can be focused on modeling the first few β_{kt}'s, and we can be much looser about modeling the remainder, imposing substantial restrictions on the corresponding parts of the model. Second, the r_{kt} are very highly cross-correlated, which is reflected in the regular shape of the fertility and mortality curves across age (see Figures 1 and 5). As shown in Bell (1988), this high cross-correlation presents statistical and possible numerical (ill-conditioning) problems in directly modeling $\underline{r}_t$. In effect, principal components attempts to remove this cross-correlation, so that transforming to $\underline{\beta}_t$ much improves the situation.

Our general approach to modeling involves the following steps: (1) We pick some small number J of principal components that provides a good approximation to the original series $\underline{r}_t$. (2) We develop a multivariate time series model for $\beta_{1t}, \cdots, \beta_{Jt}$. (3) Rather than ignore the remaining principal components, we develop univariatetime series models for them. (4) We examine whether the covariance matrix of the residuals for the complete set of principal component series can be assumed to have a simplified (block + diagonal) structure. This general approach is designed to pay closest attention to the most important principal components, while still accounting for the potential forecastibility of all them, by essentially developing a multivariate model for $\underline{\beta}_t$ with a

greatly simplified structure and a manageable number of parameters. The resulting model implies a multivariate model for $\underline{r}_t = \Lambda\underline{\beta}_t$.

To illustrate (3) let us suppose the multivariate and univariate models are of autoregressive form. Putting these models together we have

$$\begin{bmatrix} \Phi(B) & & & \\ & \phi_{J+1}(B) & & \\ & & \ddots & \\ & & & \phi_K(B) \end{bmatrix} \underline{\beta}_t = \underline{a}_t \qquad (2.1)$$

where $\Phi(B) = I - \Phi_1 B - \cdots - \Phi_p B^p$ is a J x J matrix polynomial in the backshift operator B, $\Phi_{J+1}(B), \cdots, \Phi_K(B)$ are scalar polynomials in B, and $\underline{a}_t$ is a K x 1 vector white noise series with covariance matrix Σ. There may be some deficiencies in the model (2.1) from assuming $\beta_{J+1,t}, \cdots, \beta_{Kt}$ follow univariate models and are not involved in the multivariate model for $\beta_{1t}, \cdots, \beta_{Jt}$ except possibly through contemporaneous correlation of residuals. The effect of these deficiencies should be small, however, since the contribution of $\beta_{J+1,t}, \cdots, \beta_{Kt}$ is itself small.

The simplified structure for $\Sigma = \text{Var}(\underline{a}_t)$ referred to in (4) is the following:

$$\Sigma = \begin{bmatrix} \Sigma_m & & & \\ & \sigma^2_{m+1} & & \\ & & \ddots & \\ & & & \sigma^2_K \end{bmatrix} \qquad (2.2)$$

where Σ_m is a general m x m covariance matrix for some $m \geq J$. With Σ of the form (2.2) theunivariate models for $\beta_{m+1,t}, \cdots, \beta_{Kt}$, and the multivariate model for $\beta_{1t}, \cdots, \beta_{mt}$ defined within (2.1), can be fit separately. There are two motivations for considering Σ of the form (2.2). The first is simply that a full K x K Σ involves a large number of parameters. The second is that principal components is trying to un-cross-correlate the series. While it cannot do this exactly since the series are autocorrelated, we shall see in section 3 that, for mortality, the least important principal components tend to be the least autocorrelated. Thus, it seems reasonable to investigate whether some of the less important principal component series might be not cross-correlated at all.

To select a model from among the large set of models defined by different structures for Σ as in (2.2), and with possible alternative model choices for the β_{kt}'s, we use Akaike's Information Criterion (AIC) (Akaike 1973). For multivariate AR models, AIC is defined as

$$AIC = N \log|\Sigma| + 2(np)$$

where N is the effective number of observations, $|\Sigma|$ is the determinant of the estimated residual covariance matrix Σ, and np is the number of estimated parameters. The model with the smallest AIC is preferred. When comparing two nested models, AIC is related to the likelihood ratio test statistic

$$LRT = \tilde{N} \log\frac{|\Sigma_a|}{|\Sigma_b|} = \tilde{N}(\log|\Sigma_a| - \log|\Sigma_b|)$$

where Σ_a and Σ_b are the estimated residual covariance matrices for the two models a and b, and $\tilde{N}$ is a normalizing constant that is of order N. LRT is asymptotically distributed as χ^2_v, where v is the number of parameters in model a constrained to zero in model b. Likelihood ratio tests would be difficult to use here because of the variety and number of models we compare and because some of the models we compare are not nested. One would not expect results using AIC and LRT to disagree profoundly for nested models.

We use regression indicator variables in our modeling to account for outliers in our (principal component) series. The series are investigated for the presence of outliers using univariate models essentially through the approach of Bell (1983), which is based on methodology described in more detail in Chang, Tiao, and Chen (1988). Level shifts (LS) and additive outliers (AO) are handled via regression indicator variables defined for any year t as follows:

$$AO_t^{(t_0)} = \begin{cases} 1, & \forall t = t_0 \\ 0, & \forall t \neq t_0 \end{cases}; \quad LS_t^{(t_0)} = \begin{cases} -1, & \forall t < t_0 \\ 0, & \forall t \geq t_0 \end{cases}.$$

Level shifts were defined using a (-1,0) pattern rather than the more traditional (0,1) pattern to simplify the calculation of forecasts. The regression terms used then do not affect the forecasts, so their effects can be subtracted out and need not be added back in. We subtract out the estimated outlier effects rather than including their regression terms in the multivariate models to simplify the multivariate model fitting. The effects of this type of outlier adjustment are shown in Figures 3.a,b (for the logarithm of the age 0 central death rates).

2.2 Developing Point and Interval Forecasts Suppose we have fitted a model of form (2.1) and (2.2) to data through time n, and wish to forecast the rates $\underline{r}_t$ at time t = n+l for some l > 0. Point forecasts, $\underline{\beta}^*_{n+l}$, and forecast error variance matrices $\text{Var}(\underline{\beta}_{n+l} - \underline{\beta}^*_{n+l})$, for $\underline{\beta}_{n+l}$, can be computed from the fitted model for $\underline{\beta}_t$ as described in Tiao and Box (1981). We then convert these to point forecasts and forecast error variance matrices for $\underline{r}_t$ using the following relations:

$$\underline{r}^*_{n+l} = \Lambda \underline{\beta}^*_{n+l} \qquad (2.3)$$

$$\begin{aligned} Var(\underline{r}_{n+l} - \underline{r}^*_{n+l}) &= \Lambda V_l \Lambda' \\ &= \Lambda_m V_{m,l} \Lambda_m' + \sum_{j=m+1}^{K} v^2_{j,l} \lambda_j \lambda_j' \end{aligned} \qquad (2.4)$$

where

$$V_l = \begin{bmatrix} V_{m,l} & & & \\ & v^2_{m+1,l} & & \\ & & \ddots & \\ & & & v^2_{K,l} \end{bmatrix} = Var(\underline{\beta}_{n+l} - \underline{\beta}^*_{n+l})$$

and

$$V_{m,l} = \sum_{j=1}^{l-1} \Psi_j \Sigma_m \Psi_j'$$

is the estimated m x m forecast error variance matrix in the model for $\beta_{1t}, \cdots, \beta_{mt}$ (the Ψ_j's being the psi-weight matrices for the model); $v^2_{j,l}$ is the estimated forecast error variance from the univariate model for β_{1t} for j = m+1, $\cdots$, K; Λ_m contains the first

m columns of Λ, and $\lambda_{m+1}, \cdots, \lambda_K$ are the remaining columns of Λ. Prediction intervals for the elements of r^*_{n+l} follow from the results of (2.3) and (2.4) in the usual way. In (2.4) we have ignored the contribution to the forecast error of estimating rather than knowing model parameters, though terms could be added to $V_{m,l}$ and the $v^2_{j,l}$ to allow for this (see Ansley and Kohn 1986).

If r_t is the logarithm of an original series of rates, say $r_t = \log(R_t)$ with $\log(\bullet)$ taken componentwise, then point forecasts can be obtained by inverting the transformation, $R^*_{k+l} = \exp(r^*_{k+l})$, and the corresponding forecast standard errors are in percentage terms. An approximate 95% prediction interval for $R^*_{k,n+l}$ is given by $(\exp\{r^*_{k,n+l} - 2\,[\mathrm{Var}(r_{k,n+l} - r^*_{k,n+l}))]^{1/2}\}, \exp\{r^*_{k,n+l} + 2\,[\mathrm{Var}(r_{k,n+l} - r^*_{k,n+l}))]^{1/2}\})$. An alternative point forecast is $R^{**}_{k,n+l} = \exp\{r^*_{k,n+l} + .5\,\mathrm{Var}(r_{k,n+l} - r^*_{k,n+l})\}$, which is nearer to being unbiased and having minimum mean squared error for R_{n+l} (Granger and Newbold 1976). If is some other transformation of R_t this can be dealt with on a case-by-case basis.

3. Modeling and Forecasting Central Death Rates Here we use central death rates for white women in the U.S. from 1940-87 for single years of age 0 through 4, for 5-year age groups from 5-9 through 80-84, and for 85+. This yields 22 age-specific time series of 48 observations each. The central death rates are obtained from death figures published annually by the National Center for Health Statistics (1940-1987), and from unpublished population figures compiled by the Population Division of the Census Bureau. The population figures are consistent with those in Census Bureau publications. (See U. S. Bureau of the Census 1965, 1974, 1982, 1990.) To apply the principal components approach we let $r_{kt} = \log(m_{kt})$ where k indexes the 22 age groups, and m_{kt} is the central death rate for age group k in year t.

Graphs of r_{kt} over all ages for four years are shown in Figure 1a-1d. We see a common pattern each year, with the most pronounced change over time being an increase in the magnitude of the "accident hump" at ages 15-19 each decade. Figure 2 shows r_{kt} over all years for some ages, along with point and interval forecasts developed subsequently. These graphs show the general downward trend of mortality, with some exceptions. One exception is the year 1943, when central death rates increase at all ages except age 0. This may, of course, have something to do with data problems related to U.S. entry into World War II, though we have nothing to confirm this or any other explanation. There is also an increase in central death rates at ages 15-19 for roughly the years 1965 through 1980 that we have no explanation for. There is an explanation for the erratic behavior of r_{0t} in some years, however. In years when births increased significantly through the year, r_{0t} tends to increase over the past year, since the population figures in the denominators of r_{0t} are estimates as of July 1, and so are understated in this case. The most profound example of this phenomenon occurs in 1946. The reverse can potentially occur when births decrease significantly through a year. To avoid these problems caused by data for age 0, we fit a univariate time series model to r_{0t}, performed outlier detection as in Bell (1983), and modified r_{0t} for outliers (AO's in 1943, 1946, and 1947, and an LS in 1971) before calculating principal components. The effects of this modification can be seen by comparing Figures 2.a and 2.b.

The principal component series were then examined. It was found that five principal components provided a good approximation to the original 22 time series; thus, most of the attention in model development (i.e. multivariate modeling) could be devoted to the first five principal component series. Before we proceeded with this, univariate models were identified for all the principal component series. For β_{13t} through β_{22t} there appeared to be little autocorrelation, suggesting white noise or perhaps AR(1) models for these series. The univariate models for β_{1t} through β_{12t}, along with outlier detection results, are given in Bell and Monsell (1991). Only β_{1t}, β_{2t}, and β_{3t} are differenced; the remaining principal component series are assumed to follow stationary models. The models for β_{1t} and β_{2t} allow for nonzero means in the differenced series (trend constants). Graphs of β_{1t} and β_{2t} showed steady downward movements over time, and the trend constants estimated in their models were highly significant. β_{3t} did not have a significant downward or upward trend, however. Apparently, the trend constants in the models for β_{1t} and β_{2t} are what is needed to capture the overall downward trend in mortality over time evident in all the graphs of Figure 2.

The effects of the outliers detected in $\beta_{1t},\ldots,\beta_{12t}$ were subtracted out before proceeding to the multivariate modeling. An additive outlier (AO) was detected in β_{6t} for t = 1987, the last year of our data set. One effect of adjusting for this outlier is that forecasts of β_{6t} will show a discontinuity from $\beta_{6,1987}$, and be more in conformance with values of β_{6t} in years prior to 1987. It will be seen later that this produces similar discontinuities from the last data point in forecasts of log central death rates r_{kt} at many age groups k. It is important to realize that when the last observation of a time series appears to be an outlier, it is initially impossible to tell (see Bell 1983) whether this is a temporary aberration (AO) or a permanent change (LS). As the characterization of an outlier at or near the end of a series has important effects on the forecasts, it is important to monitor succeeding observations as they are obtained, to reassess the outlier in light of new information.

We next developed a multivariate model for $Z_t = (\nabla\beta_{1t}, \nabla\beta_{2t}, \nabla\beta_{3t}, \beta_{4t}, \beta_{5t})'$. Examining matrices of auto- and cross-correlations, and also stepwise autoregressive fits, as in Tiao and Box (1981), suggested either a moving average or autoregressive model of order 1. Our attempts to fit a multivariate MA(1) model eventually resulted in a noninvertible model, so we dropped this in favor of the multivariate AR(1) model. Starting with an AR(1) model with Φ_1 a full 5 x 5 matrix, estimating this, setting elements of Φ_1 with t-statistics less than 2 in absolute value to 0, reestimating and continuing this process, we eventually obtained the following:

$$\Phi_1 = \begin{bmatrix} .73 & & & & \\ .15 & 0 & & & \\ & & 0 & & \\ & & & .82 & \\ & & & & .79 \end{bmatrix} \quad se(\Phi_1) = \begin{bmatrix} .09 & & & & \\ .03 & \cdot & & & \\ & & \cdot & & \\ & & & .06 & \\ & & & & .08 \end{bmatrix}$$

We then combined this multivariate model for Z_t with the univariate models for $\beta_{6t},\ldots,\beta_{12t}$ given in Bell and Monsell

(1991) and with either AR(1) or white noise univariate models for $\beta_{13t},\ldots,\beta_{22t}$. We estimated these models of form (2.1) for β_t with Σ having the form (2.2) for various values of m, here ranging from 5 to 22. The results of the AIC analysis are given in Bell and Monsell (1991). The AR(1) models for the last components ($\beta_{13t},\ldots,\beta_{22t}$) are favored over the white noise models. AIC picks a model with m=6, though the AIC values do not increase uniformly with m; a model with m=11 would be the second choice. In what follows we use the model with m=6.

This model for β_t implies a model for $r_t = \Lambda\beta_t$ and then point and interval forecasts from 1987 for r_t are developed as discussed in section 2.2. These are shown in Figures 1.e-1.h for all ages for a few years. Figure 2 shows corresponding point and interval forecasts of $m_{kt} = \exp(r_{kt})$ for a few ages for the years 1988 through 2010. We notice the following. (1) The historical shape of the mortality curve over age is captured in the predictions shown in Figure 1. (2) There appears to be relatively less uncertainty in forecasting central death rates than in forecasting fertility rates, see Bell and Monsell (1991). (3) The uncertainty in forecasting $r_{kt} = \log(m_{kt})$ is largest roughly at ages 1 through 40, with the width of the forecast intervals narrowing at age 0 or as age increases above 40. However, because the point forecasts at ages 1 through 40 are also low, the most uncertainty in forecasting $m_{kt} = \exp(r_{kt})$ is actually at the other ages. In fact, we do not show plots of forecasts of m_{kt} over age for given years because the forecast intervals in such plots essentially show up only at the advanced ages. (Notice the different plots in Figure 2 are on different scales.) (4) The forecasts at several ages in Figure 2 show slight discontinuities from the last data point, due to the adjustment for the outlier in β_{6t} for 1987 noted earlier. (5) Figure 2.h shows that at ages 70-74 central death rates decreased fairly steadily to about 1980, and have remained fairly flat since then. The forecasts at ages 70-74, however, show a downward progression of central death rates consistent with the historical data prior to 1980 but inconsistent with the data since that time. Similar results were obtained for age groups 65-69, 75-79, 80-84, and 85+. This forecast of improved mortality at older ages derives from the forecasted general improvement of mortality, which in turn results from the forecasted declines of β_{1t} and β_{2t} mentioned earlier. Opinions may, of course, differ as to whether central death rates at older ages (here, for white females) will improve in the future, particularly since mortality improvement in recent years at these ages has been slight. Those who do not expect much future improvement in mortality at older ages could treat these ages separately in forecasting. Principal components might be applied twice: first for age groups below 65 and then for the 65 and over age groups. One might then try to link the resulting two multivariate models in some fashion, say through a joint residual covariance matrix.

4. Conclusions Projecting age-specific mortality is a forecasting problem of high dimension. A multivariate time series model for the full set of age-specific rates can be useful in producing point and, especially, interval forecasts. Unfortunately, because of the high dimension <u>and</u> the high correlation between age-specific rates (as evidenced by the smooth shape of the rates across age), direct attempts at time series modeling and forecasting of these series are not likely to be successful. In this paper we have shown how this problem can be addressed by computing principal components of the data, developing a multivariate time series model of greatly simplified structure for the full set of principal components (which are much more amenable to multivariate time series modeling than the original rates), using this model to forecast the principal components, and translating these results into point and interval forecasts of the age-specific rates. The interval forecasts derive from the conditional distribution of future data given past data implied by the model. An important topic for future work is to use this conditional distribution to produce prediction intervals for births and deaths, rather than just the rates, and ultimately to produce prediction intervals for the age-specific population figures.

5. Disclaimer This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

References

Akaike, H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle," ***2nd International Symposium on Information Theory***, ed. B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado, 267-281.

Ansley, C. F. and Kohn, R. (1986) "Prediction Mean Squared Error for State Space Models with Estimated Parameters," ***Biometrika***, 73, 467-473.

Bell, W. R. (1983) "A Computer Program for Detecting Outliers in Time Series," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 634-639.

_____ (1988) "Time Series Models for Forecasting Fertility," paper presented at the International Workshop on National Population Projections in Industralized Countries, October 1988, Voorburg, The Netherlands.

Bell, W. R. and Monsell, B. C. (1991) "Using Principal Components in Time Series Modelling and Forecasting of Age-Specific Demographic Data," to appear in the SRD Research Report Series.

Box, G.E.P. and Jenkins, G. M. (1970), ***Time Series Analysis: Forecasting and Control***, San Francisco: Holden Day.

Bozik, J. E. and Bell, W. R. (1987) "Forecasting Age Specific Fertility Using Principal Components," SRD Research Report 87/19, Statistical Research Division, Bureau of the Census, Washington, D.C.

Chang, I., Tiao, G. C., and Chen, C. (1988) "Estimation of Time Series Parameters in the Presence of Outliers," ***Technometrics***, 30, 193-204.

Granger, C.W.J. and Newbold, P. (1976) "Forecasting Transformed Series," ***Journal of the Royal Statistical Society, Series B***, 38, 189-203.

National Center for Health Statistics (1940-1987) ***Vital Statistics of the United States, Volume II — Mortality***, Washington, D.C.: U.S. Government Printing Office.

Tiao, G. C. and Box, G.E.P. (1981) "Modeling Multiple Time Series With Applications," ***Journal of the American Statistical Association***, 76, 802-816.

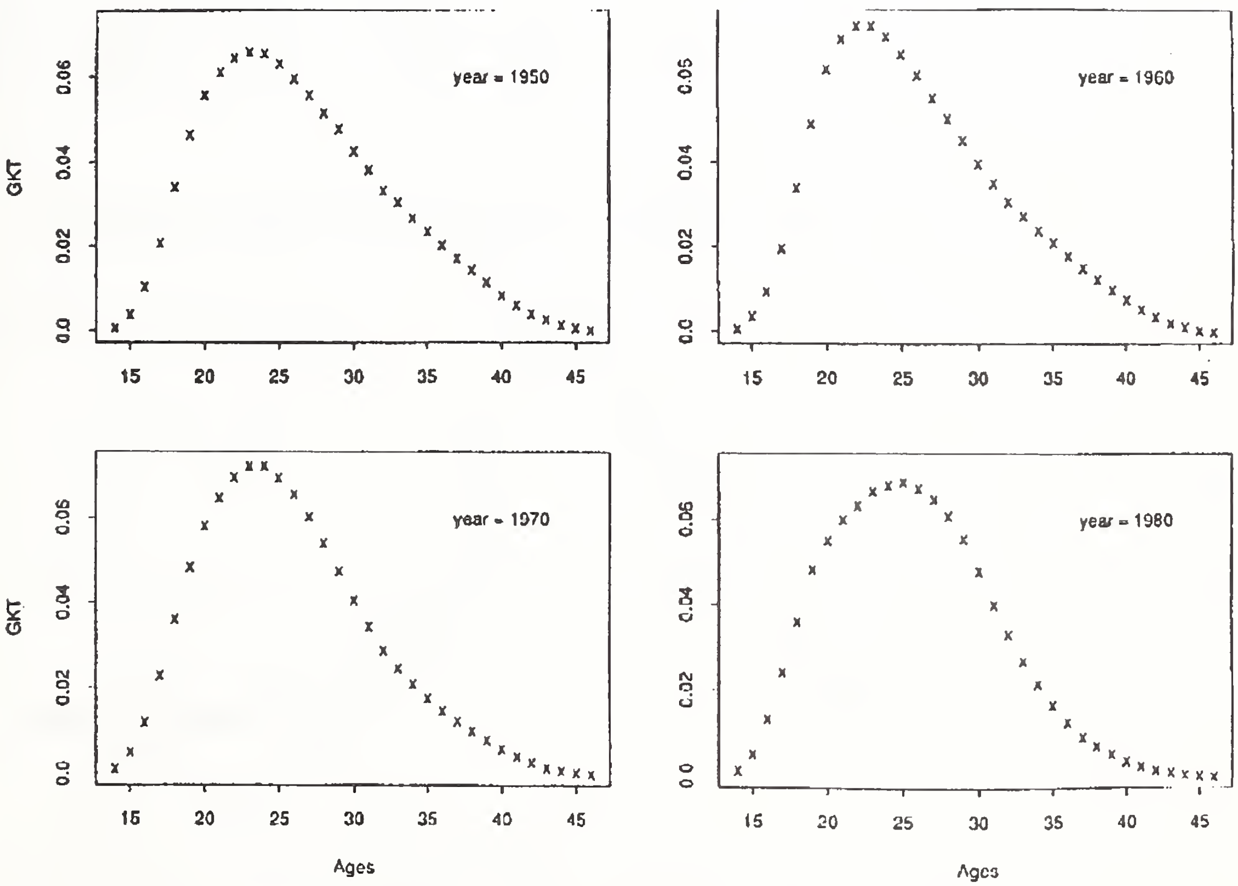
U.S. Bureau of the Census (1965), "Estimates of the Population of the United States by Single Year of Age, Color, and Sex, 1900-1959", Current Population Reports, Series P-25, No. 311, Washington, D.C.: U.S. Government Printing Office.

_____ (1974), "Estimates of the Population of the United States by Age, Sex, and Race, April 1, 1960 to July 1, 1973", Current Population Reports, Series P-25, No. 519, Washington, D.C.: U.S. Government Printing Office.

_____ (1982), "Preliminary Estimates of the Population of the United States by Age, Sex, and Race, 1970-1981", Current Population Reports, Series P-25, No. 917, Washington, D.C.: U.S. Government Printing Office.

_____ (1990), "United States Population Estimates, by Age, Sex, Race and Hispanic Origin, 1980-1988", Current Population Reports, Series P-25, No. 1045, Washington, D.C.: U.S. Government Printing Office.

Figure 1 : Age-Specific Fertility



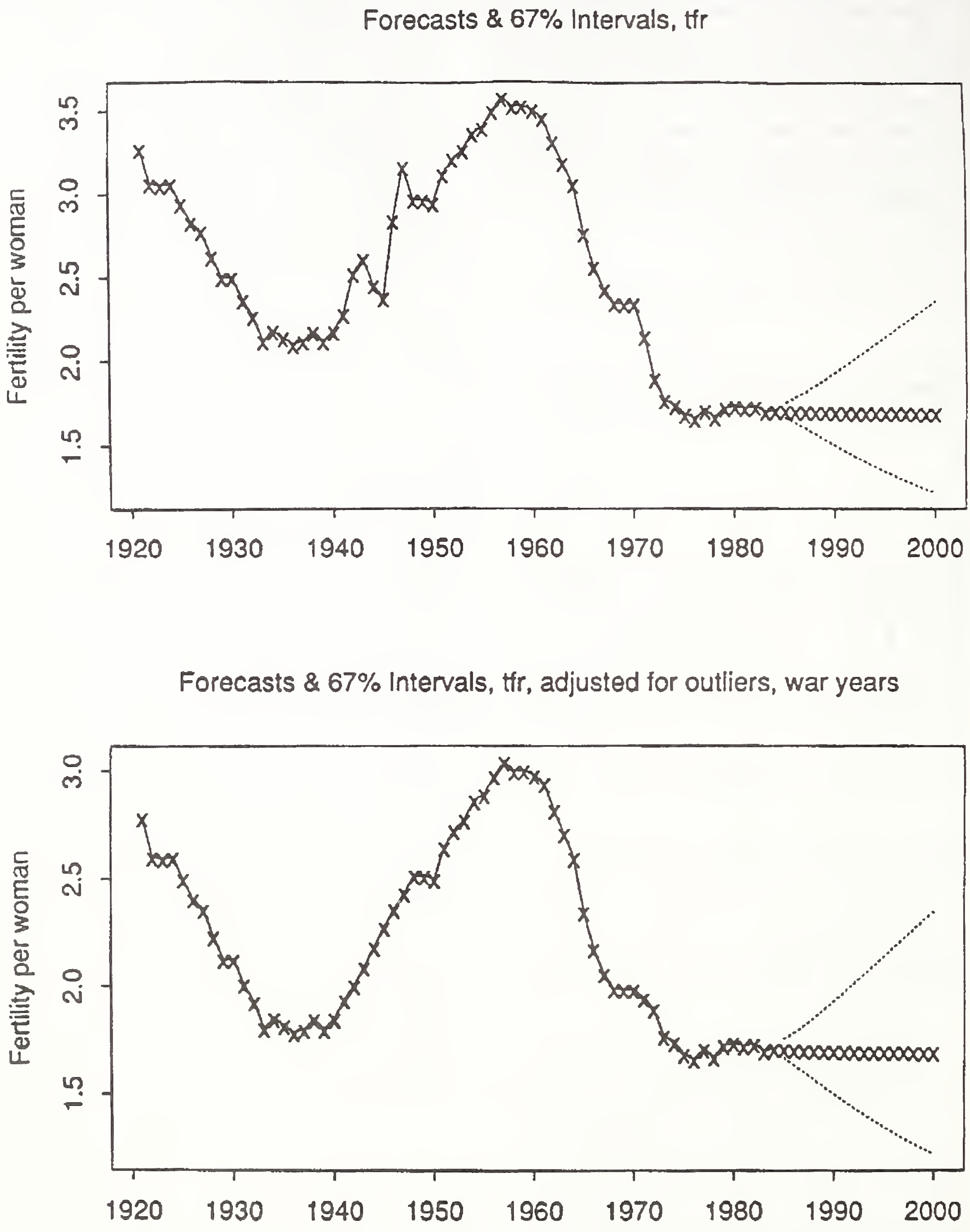

Forecasts & 67% Intervals, tfr
Fertility per woman
1.5
2.0
2.5
3.0
3.5
1920
1930
1940
1950
1960
1970
1980
1990
2000
Forecasts & 67% Intervals, tfr, adjusted for outliers, war years
Fertility per woman
1.5
2.0
2.5
3.0
1920
1930
1940
1950
1960
1970
1980
1990
2000

Figure 2

Forecasts & 67% Intervals

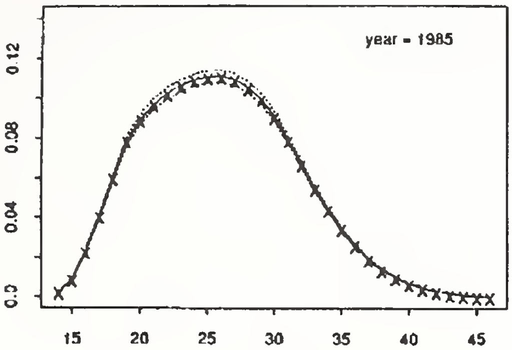


Figure 3.a

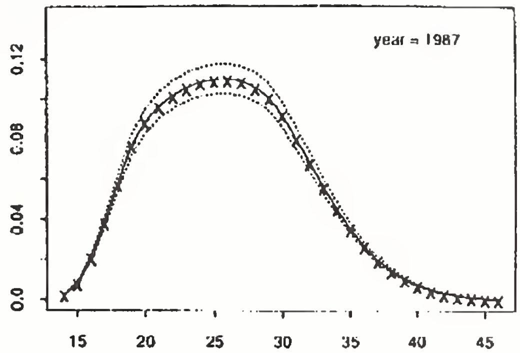


Figure 3.b

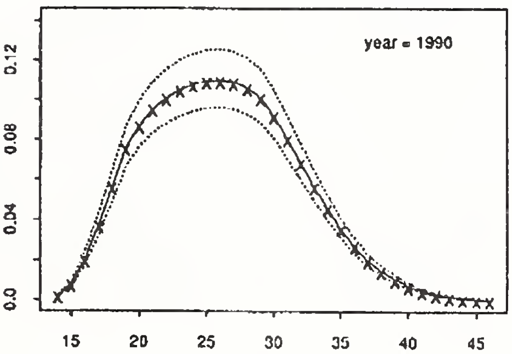


Figure 3.c

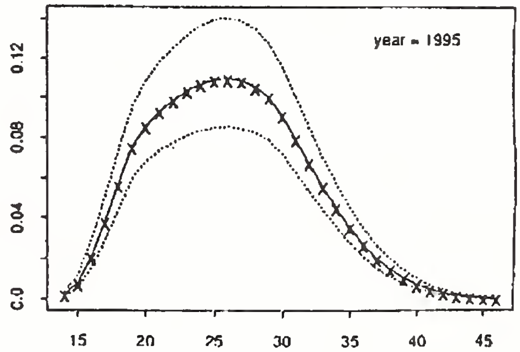


Figure 3.d

Forecasts & 67% Intervals

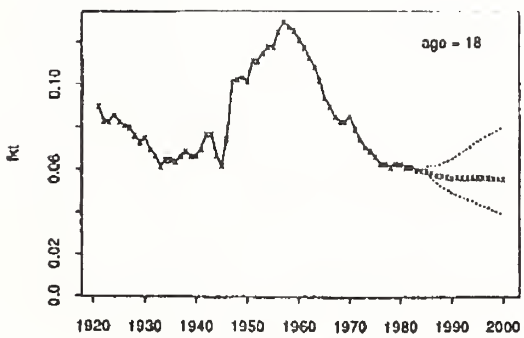


Figure 4.a

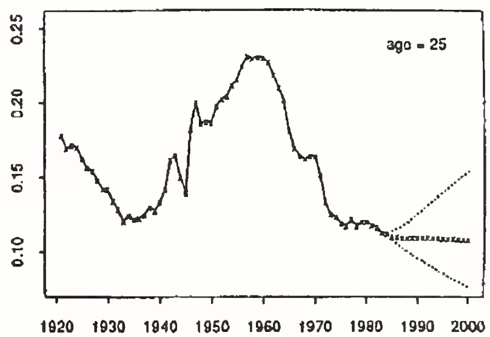


Figure 4.b

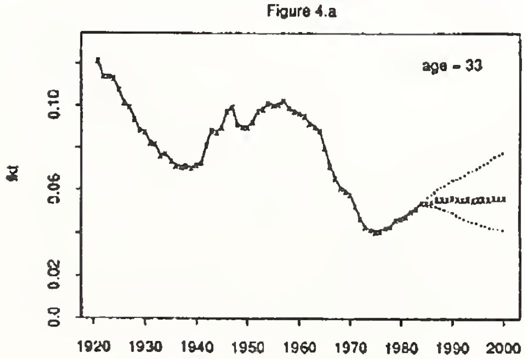


Figure 4.c

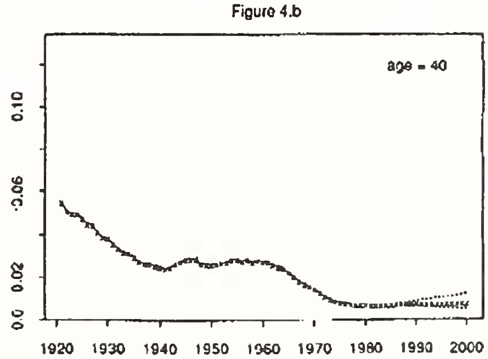
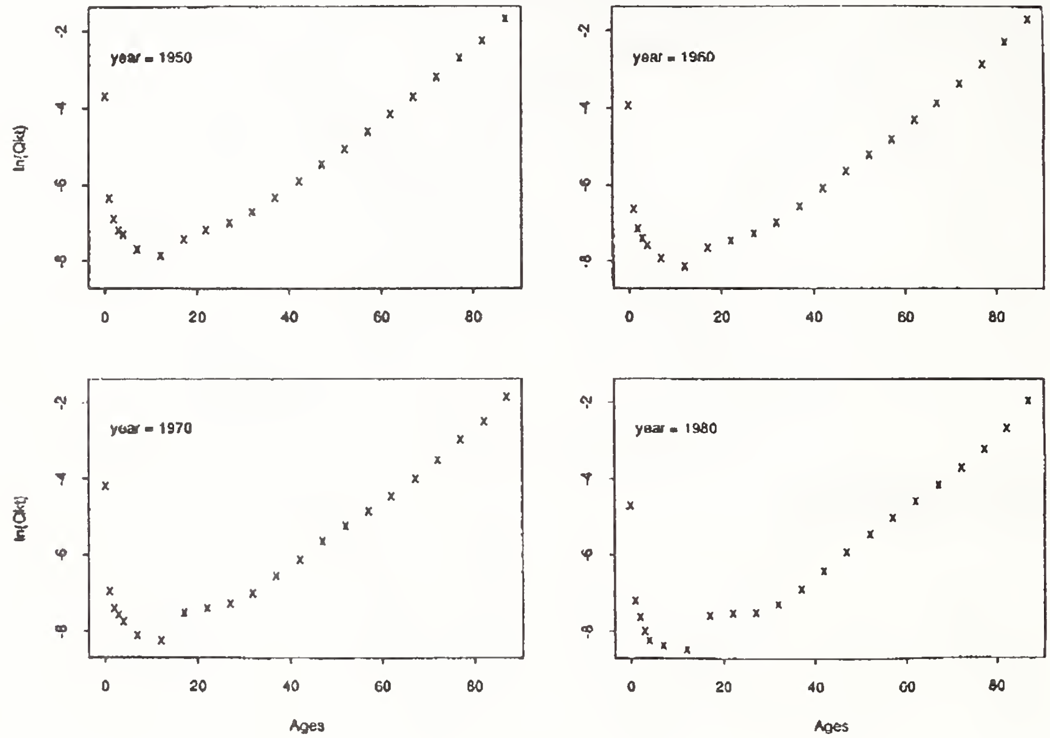


Figure 4.d

Figure 5 : Log of White Female Age-Specific Mortality



Forecasts & 95% Intervals

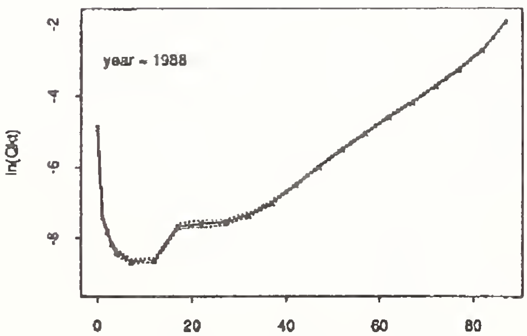


Figure 6.a

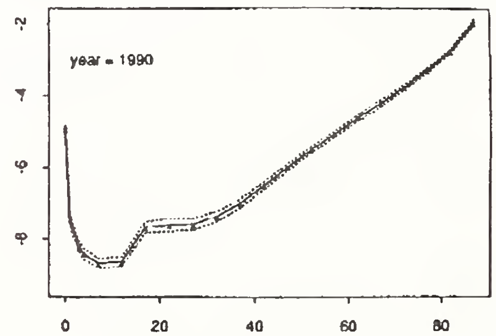


Figure 6.b

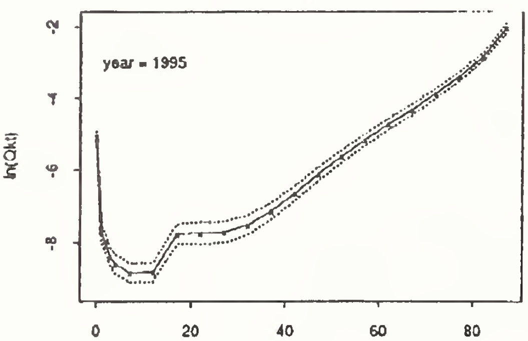


Figure 6.c

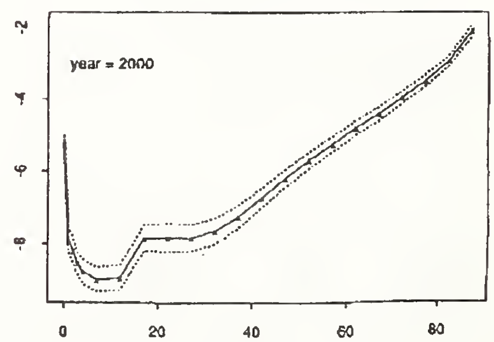


Figure 6.d

Forecasts & 95% Intervals

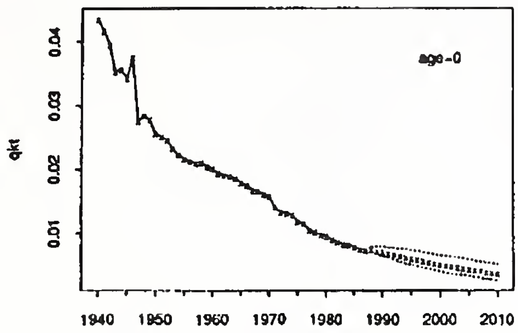


Figure 7.a

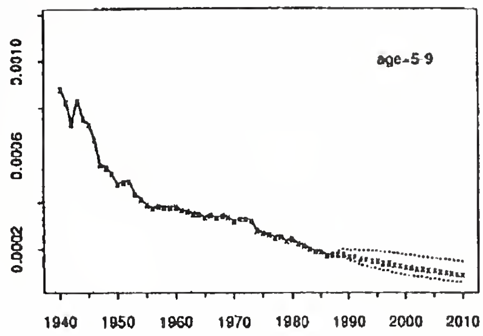


Figure 7.c

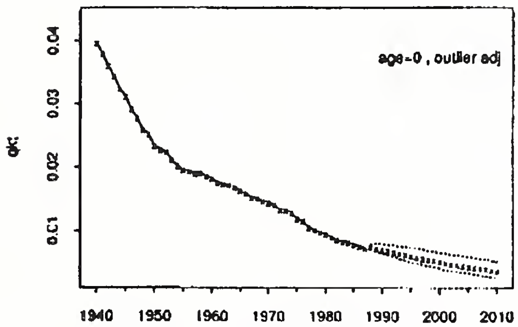


Figure 7.b

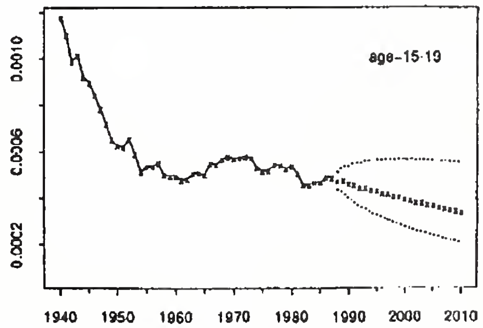


Figure 7.d

Forecasts & 95% Intervals

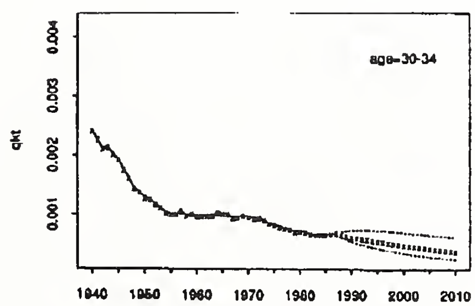


Figure 7.e

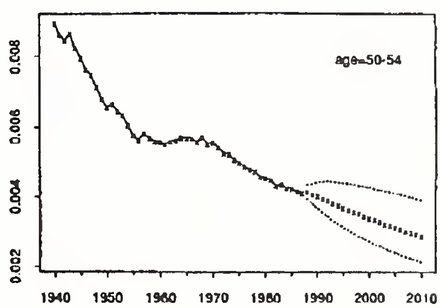


Figure 7.g

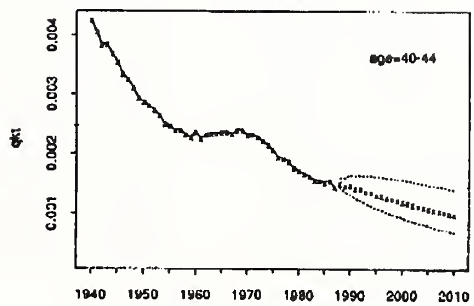


Figure 7.f

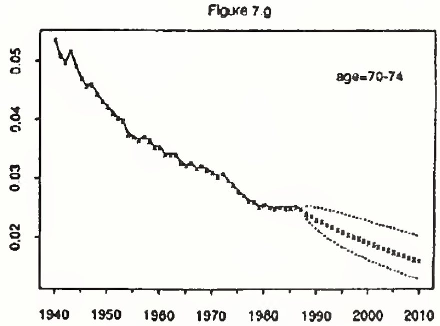


Figure 7.h

TABLE 1

Univariate Models and Outliers Identified for $LTFR_t$, β_{1t}, $\cdots$, β_{12t} for Fertility Rates

COMPONENT NUMBER	ARIMA MODEL	OUTLIERS IDENTIFIED
LTFR	(3 1 0)	AO + LS in 1971
1	(3 1 0)	AO in 1984, LS + AO in 1974
2	(2 1 0)	AO in 1983, LS in 1974
3	(1 1 0)	None
4	(1 1 0)	AO in 1975, LS in 1972, 1980
5	(4 0 0)	AO in 1964, 1968, 1975
6	(5 0 0)	None
7	(1 0 0)	AO in 1976, 1980, 1984
8	(3 0 0)	AO in 1980
9	(1 0 0)	—
10	(1 0 0)	—
11	(3 0 0)	—
12	(1 0 0)	—

Note: Outliers were not identified for β_{9t} , $\cdots$, β_{12t} .

TABLE 2

Comparing Models for LTFR and Weighted Principal Components of Relative Fertility Rates (Adjusted for Outliers and Effects of World War II)

		Model for β_{13t}, $\cdots$, β_{33t}			
		White Noise		AR(1)	
No. Jointly Modeled Series (m)	No. Uncorr. Series (34–m)	$\log\|\Sigma\|$	AIC	$\log\|\Sigma\|$	AIC
34	0	–288.3	–15,732.1	–288.3	–15,732.0
29	5	–286.0	–15,848.9	–286.1	–15,849.5
25	9	–283.4	–15,905.1	–284.2	–15,945.5
21	13	–281.8	–15,985.0	–282.7	–16,038.8
17	17	–280.6	–16,095.3	–281.6	–16,124.1
16	18	–280.3	–16,109.8	–281.4	–16,139.4
15	19	–279.8	–16,114.3	–280.9	–16,141.9
14	20	–279.6	–16,131.8	–280.7	–16,158.5
13	21	–279.3	–16,142.2	–280.5	–16,174.6
12	22	–279.2	–16,159.7	–280.4	–16,192.1
11	23	–279.0	–16,169.6	–280.2	–16,202.1
10	24	–278.7	–16,173.6	–280.0	–16,206.0
9	25	–278.5	–16,179.7	–279.8	–16,212.1 †
8	26	–278.3	–16,184.3	–279.6	–16,216.8 *
7	27	–277.7	–16,161.6	–278.9	–16,194.1
6	28	–277.6	–16,169.7	–278.9	–16,202.1
5	29	–277.5	–16,170.0	–278.7	–16,202.4

* Model favored by AIC
† Model used

TABLE 3

Univariate Models and Outliers Identified for $\beta_{1t}, \cdots, \beta_{12t}$ for Mortality Rates

COMPONENT NUMBER	ARIMA MODEL	OUTLIERS IDENTIFIED
1	μ + (1 1 0)	AO in 1942
2	μ + (0 1 0)	—
3	(0 1 0)	LS in 1948
4	(1 0 0)	—
5	(1 0 0)	AO in 1947, 1952
6	μ + (0 0 0)	AO in 1951, 1987
7	(2 0 0)	LS in 1945, 1951
8	(3 0 0)	AO in 1956
9	μ + (0 0 0)	—
10	(1 0 0)	—
11	(1 0 0)	AO in 1948
12	(3 0 0)	—

Note: μ denotes a nonzero mean in the model. This is for the differenced series for β_{1t} and β_{2t}, i.e. it is a trend constant there.

TABLE 4

Comparing Models for Principal Components of Mortality Rates (adjusted for outliers)

		Model for $\beta_{13t}, \cdots, \beta_{22t}$			
		White Noise		AR(1)	
No. Jointly Modeled Series (m)	No. Uncorr. Series (22–m)	$\log\|\Sigma\|$	AIC	$\log\|\Sigma\|$	AIC
22	0	–177.9	–7,256.5	–179.0	–7,308.4
21	1	–177.7	–7,289.0	–178.6	–7,335.2
20	2	–177.3	–7,314.1	–178.3	–7,361.7
19	3	–177.1	–7,346.1	–177.8	–7,380.2
18	4	–176.8	–7,372.8	–177.4	–7,402.1
17	5	–176.4	–7,391.2	–176.9	–7,414.1
16	6	–175.9	–7,403.4	–176.4	–7,426.0
15	7	–175.6	–7,420.0	–176.1	–7,445.7
14	8	–175.4	–7,440.7	–175.9	–7,466.2
13	9	–175.1	–7,461.5	–175.6	–7,482.5
12	10	–174.8	–7,473.0	–175.3	–7,493.3
11	11	–174.6	–7,492.5	–175.1	–7,512.7
10	12	–173.9	–7,483.7	–174.4	–7,503.9
9	13	–173.5	–7,484.8	–173.9	–7,505.0
8	14	–173.0	–7,482.5	–173.5	–7,502.7
7	15	–172.7	–7,490.5	–173.2	–7,510.7
6	16	–172.5	–7,496.5	–173.0	–7,516.7 *
5	17	–172.0	–7,486.3	–172.5	–7,506.5

* Model favored by AIC

Developing Long-Term Energy Projections for the National Energy Strategy

Arthur Rypinski, U. S. Energy Information Administration

On July 26, 1989, the President of the United States directed the Secretary of Energy to develop a comprehensive National Energy Strategy (or NES, as it was frequently abbreviated). However, between its beginnings in the summer of 1989, and the issuance of the "First Edition" of the NES in the spring of 1991, a great deal of work was done.

This essay will focus largely on a synopsis of the most important results of the Energy Information Administration's (EIA) modeling effort, with only the briefest bow in the direction of a discussion of the modeling and policy formulation process.

Some Points About Process

The development of the National Energy Strategy went through three phases:

- eliciting public comment, from July, 1989 through April, 1990;
- development of policy options and interagency consultation, from April, 1990 to February, 1991;
- Legislative and Executive Branch implementation, from March, 1991 to the present.

My agency, the Energy Information Administration, was primarily involved with the second phase. Our task was to develop a hypothesized energy future against which various policy options could be examined. As part of this task, we prepared a set of seven NES-related service reports listed in the Appendix. The EIA service reports do not, however, trace the potential impacts of the actual NES legislative and administrative initiatives as developed during the winter of 1990-91. These initiatives were developed by interagency working groups. During the period of actual policy formulation, EIA acted, in effect, as a consultant, answering questions posed by the Department, particularly in predicting the macroeconomic impact of various proposed policies.

Following interagency consultation, the DOE developed a set of "final" projections of energy supply and demand, and a set of "final" estimates of the impact of the proposed policies, which were published in the February 1991 NES report.

However, the published EIA report, the final NES report, and other EIA and DOE published projections share a strong family resemblance. As often the case, while there innumerable differences in detail, the broad features of the forecast are very similar.

These projections contain an enormous amount of detail. However, the major features, the "big picture" if you will, can comfortably be inscribed on a three by five card. The big picture is shared, not only by the EIA report and the "final" NES projections, but also by a number of other recent EIA and DOE reports. It comes close to being DOE's "corporate view" of America's energy future, and, as such, has an interest beyond its mere accuracy as a forecast.

Results of the Analysis: The Baseline Scenario

The detailed discussion of factors which could affect future energy demand is documented in the EIA service reports. Rather than describing how many quads are used for what in the year 2000, it would be more useful, and certainly more interesting, to take a page from the book of some of my colleagues who do "scenario analysis" and describe some common features of possible energy futures using a much broader brush, describing first, a baseline scenario, and then some alternative scenarios. This approach also has the advantage of describing broad trends that are common to much of the NES analysis.

The baseline scenario is a 40-year forecast of U.S. energy production, consumption, and prices. The key assumptions on GNP growth were provided by the Council of Economic Advisors, and the assumptions on world oil prices were generated by EIA's Oil Market Simulation model. The technology assumptions that used in these reports are often quite optimistic. In 2030, for example, coal-fired fuel cells produce electricity with 68 percent efficiency, the average new car gets 42 miles per gallon, and the average efficiency of the commercial aircraft fleet rises by 30 percent. The key results are:

U.S. energy consumption will continue to grow, though much more slowly than the U.S. economy, as energy-saving technologies continue to filter into the capital stock. Electricity consumption grows more rapidly than overall energy consumption.

Energy prices will rise after 2000. Energy prices in the future will be higher than today. Oil and gas prices will more than double, coal prices will rise about 50 percent, while average real electricity prices will be about 15 percent higher in 2020 than today (Figure 2).

Oil prices are driven by assumptions about the world oil market, and natural gas and coal prices by assumptions about relative resource availability.

Electricity prices stay relatively low because of the abundance of coal coupled to increasingly advanced electric power generation technologies.

The United States will import more oil. U.S. oil production will continue to decline, and oil imports will rise steadily. Oil imports will rise from 44 percent of U.S. oil consumption today to over 80 percent by 2030 (Figure 3). On the other hand, if one makes the assumption that oil prices will stay low forever, then oil import volumes are even higher. Higher oil prices induce additional production from a number of "high cost" sources, such as enhanced oil recovery, and, late in the scenario, from coal liquids. However, these sources build up too slowly to offset increasing consumption.

Natural gas supplies are limited. Low cost natural gas is in limited supply, and gas prices will rise rapidly, relative to other energy prices, after 2000. (See Figure 4). During the 1990's, domestic gas production rises rapidly. After 2000, much higher gas prices induce both pipeline and LNG imports, as well as the transport of Alaska North Slope gas to the "Lower 48." These additional sources of gas, however, do not offset the decline of conventional gas production in the continental United States.

Coal use will rise. Coal resources are less costly and relatively abundant, and will be increasingly used as oil and gas prices rise. Coal fuel sources will combine with advanced power generation technologies to produce relatively low-cost electricity. Advanced coal technologies such as fluidized bed combustion are also inherently "clean." Thus, thsese technologies should also reduce emissions of sulfur oxides and nitrogen oxides, which contribute to urban smog and acid rain.

Coal also gains market share because no new nuclear power plants are assumed to be licensed, and existing plants are phased out after 2010. (Figure 5).

Emissions of "greenhouse gases" will rise. Annual energy-related emissions of carbon dioxide and methane will rise more than 80 percent by 2030 (Figure 6). On the other hand, emission trends of controlled pollutants are mixed: sulfur oxide

emissions will gradually decline, while nitrogen oxide emissions will tend to rise slightly.

The reports did not forecast emissions of volatile organic compounds or carbon monoxide associated with urban smog.

Results of the Analysis: Scenario Results

We then put together scenarios aimed at seeing how far technological innovations and regulatory changes might go in addressing some of the concerns that motivated that NES. The EIA condensed the seven scenarios originally requested into three: a "global climate change" scenario, a "concern about vulnerability" scenario, and a "reduced energy cost" scenario.

Global Climate Change. This scenario presumes public interest in reducing U.S. emissions of carbon, presumably as part of a coordinated global effort. We assume extensive implementation of end-use energy conservation technology, relaxation of constraints on importing natural gas, and lower costs for renewable energy. These assumptions result in energy consumption that is significantly lower than in the reference case, especially after 2000, as economic growth slows.

However, carbon emissions stabilize only after 2010, and only with the resumption of large-scale construction of nuclear power plants. Without nuclear power, carbon emissions continue to rise. They do so because the fuel of the future is coal. Coal is cheap and plentiful, the technology to burn coal is assumed to improve steadily, and consequently, coal and nuclear compete for the long-run baseload power generation market.

Concern About Energy Vulnerability. The second scenario presumes that one is interested in reducing vulnerability to foreign oil imports. As in the Global Climate Change Case, this scenario assumes extensive implementation of end-use energy conservation technology. It also assumes that oil companies resume exploration in currently restricted areas of the United States, and that oil and gas exploration and production technology improves significantly. It also assumes lower cost coal liquefaction facilities, and a breakthrough in fermentation technology that permits us to cheaply produce methanol from "fast rotation woody crops;" that is, from trees. Finally, it assumes that new nuclear orders are possible after 2000, though this assumption turns out to be irrelevant to the scenario results.

So, what does one learn? It is possible to stabilize oil imports into the United States. (See Figure 7). However, this stabilization is caused by increased production of high-cost fuels: coal liquids and other synthetics, boosted by imported methanol. Conservation is also essential to this outcome. However, only relatively high cost sources are available to offset declining production of conventional crude oil. Thus, the high oil prices built into the baseline scenario are also essential to this outcome. Many of these high cost fuels are also high carbon producers. The extensive use of coal liquids loses the emissions reductions gained through nuclear power.

Reducing Energy Costs. Finally, the report considered how technology improvements might reduce energy costs. This scenario assumed advanced electric power generation technologies (including nuclear power), lower cost renewable energy technologies, lower cost coal liquefaction, and improved conventional petroleum technology. The scenario results indicate that it is difficult to use technology to achieve major reductions in energy costs in the United States, beyond those already built into the baseline case (Figure 8). Partly, this is because the technologies built into the baseline case are already quite advanced. It will be challenging to achieve the levels of electricity prices already built into the baseline case.

The U.S. energy system is flexible, and contains many technologies that are close competitors. The jump from the second best technology to the best technology is often not a large one.

There is tremendous inertia in the existing capital stock. The infiltration of advanced technology at the margin has little immediate effect on the overall performance of the energy system. In virtually all of the scenarios examined, advanced technologies have little impact during the 1990's, even if they are available immediately. Many of the desirable effects of scenarios (stabilization of oil imports and carbon emissions) occur only after 2015. This is largely due to slow turnover of existing capital stock.

Further, many energy prices are set on world markets rather than in the United States, and actions taken in the United States have only an indirect effect on world prices.

The technologies modeled here have only indirect effects on scarcities of domestic low-cost oil and gas resources. Many technologies are inherently high cost technologies that will only penetrate if energy prices remain high by historical standards. Therefore, they can do no more than put a ceiling on domestic energy prices.

Concluding Remarks: What Matters

One of the most useful functions of running large-scale models is that it gives one a sense of what assumptions are most important to one's results, and which results are particularly robust. It may also be useful to share some of the ideas that emerged from the analytical stew.

- The energy system has tremendous inertia built into hundreds of billions of dollars in existing capital. This capital stock turns over very slowly. Much of the power generation capacity we will be using in 2000 is already operational today. If the economy grows more slowly than the projections used in this report, capital turnover will be even slower. Therefore, policies aimed at increasing energy efficiency, whether on the supply side or the demand side, do not show immediate results, but only gradually over long periods of time.
- World oil prices are not as important as one would think. In striking contrast to the 1970's oil has largely retreated from power generation, industrial fuel, and home heating markets in the past decade. We expect this trend to continue at projected oil prices. Consequently, the impact of higher oil prices is concentrated in the transportation market, above all, in consumer gasoline use. Oil prices, if they remain high, have, therefore, become "de-coupled" from other energy prices. On the other hand, low oil prices induce a torrent of oil imports in the model: 14 million barrels per day in 2010 in the baseline case, 18 million barrels per day if oil prices remain below $30 per barrel. It is difficult to imagine the U. S. energy system absorbing such high levels of oil imports without higher environmental, transport and refining/marketing costs, even leaving aside possible political repercussions.
- Natural Gas is the Key "Swing Fuel." We have seen the future, and its name is coal. This is fine from the point of view of cost and energy security, but it gets a bit sticky on environmental grounds. On the other hand, the technology of natural gas power generation has advanced enormously in recent years, largely due to defense-related research and development on jet engines. A U.S. manufacturer is now selling simple cycle gas turbine generators with a 41-percent efficiency for about $300 per kilowatt. Combined cycle plants with efficiencies greater than 50 percent are also available, with capital costs that are still less than half that of coal-fired

plants. If gas prices remain low, this is the wave of the future. However, our estimate, based upon studies by the U.S. Geological Survey, the Potential Gas Committee, and others, is that the resource base won't sustain the huge increases in gas consumption needed to compete seriously with coal. Rather, gas prices are expected to rise sharply after 2000, as low cost resources are depleted, and we start using much larger high cost gas resources.

- Reducing the growth of energy consumption is a key to making other problems tractable. This is an outcome of all of our scenario analysis. It is easier, however, to assert that reducing energy consumption is desirable than it is to design policy instruments that will actually reduce energy consumption.

U.S. Energy Information Administration Service Reports Related to the National Energy Strategy

All of these reports are available through the U.S. Government Printing Office, the National Technical Information Service (NTIS) or by calling the National Energy Information Center at (202) 586-8800.

U.S. Energy Information Administration, *Improving Technology: Modeling Energy Futures for the National Energy Strategy* (SR/NES/90-01) (December, 1990).

____________, *Energy Conservation and Conservation Potential: Supporting Analysis for the National Energy Strategy* (SR/NES/90-02) (December, 1990).

____________, *Electricity Supply: Supporting Analysis for the National Energy Strategy*, (SR/NES/90-03) (December, 1990).

____________, *Renewable Energy Excursion: Supporting Analysis for the National Energy Strategy*, (SR/NES/90-04) (December, 1990).

____________, *The Domestic Oil and Gas Recoverable Resource Base: Supporting Analysis for the National Energy Strategy*, (SR/NES/90-05) (December, 1990).

____________, *The Outlook for Natural Gas Imports: Supporting Analysis for the National Energy Strategy*, (SR/NES/90-06) (December, 1990).

____________, *The Potential for Coal Liquefaction: Supporting Analyses for the National Energy Strategy*, (SR/NES/90-06) (December, 1990).

____________, *Studies of Energy Taxes*, (SR/EMEU/91-02) (April, 1991).

Figure 1. Energy Prices: New Technologies (Dollars Per Million Btu)

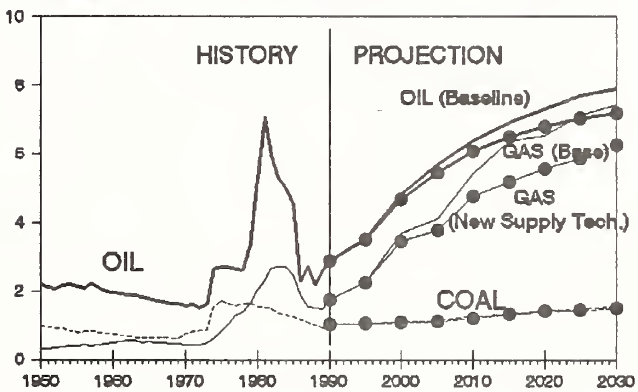


Figure 2. Energy Consumption Indices 1990=100

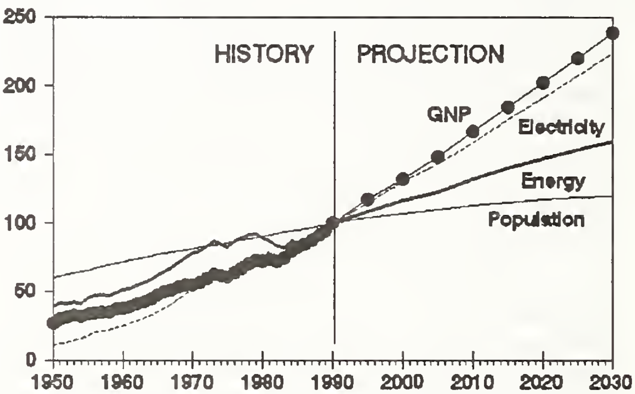


Figure 3. U.S. Energy Consumption (Quadrillion Btu)

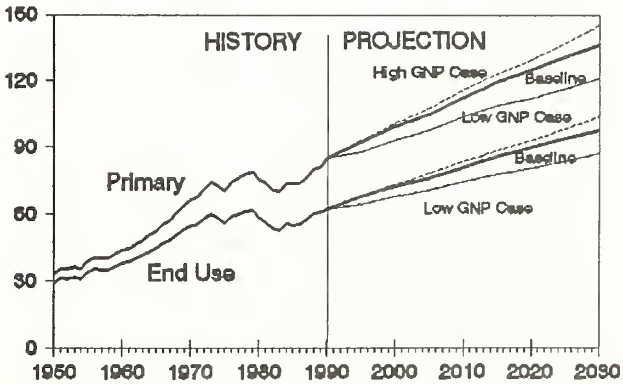


Figure 4. U.S. Energy Prices (Dollars Per Million Btu)

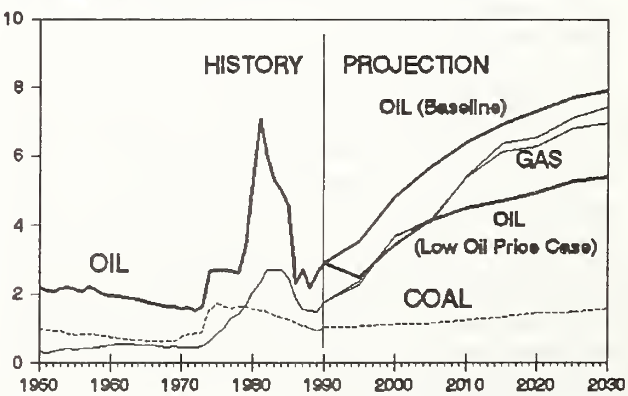


Figure 5. U.S. Energy Consumption by Fuel Type
(Quadrillion Btu)

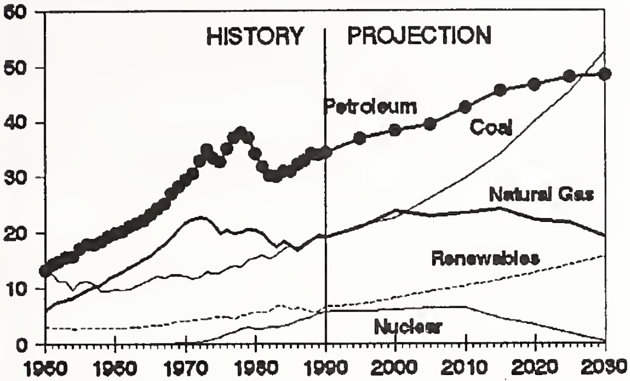


Figure 6. Sources of Natural Gas Consumption
(Quadrillion Btu)

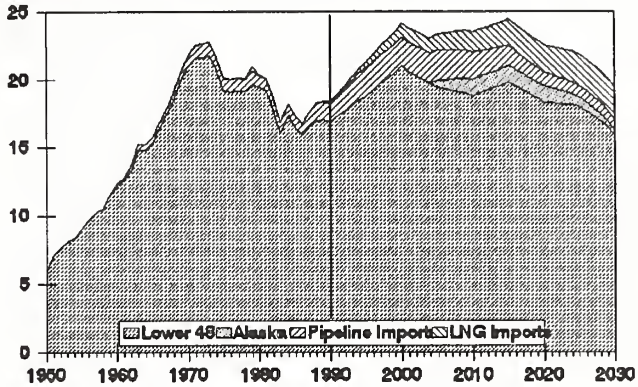


Figure 7. U.S. Energy-Related Carbon Emissions
(Million Metric Tons)

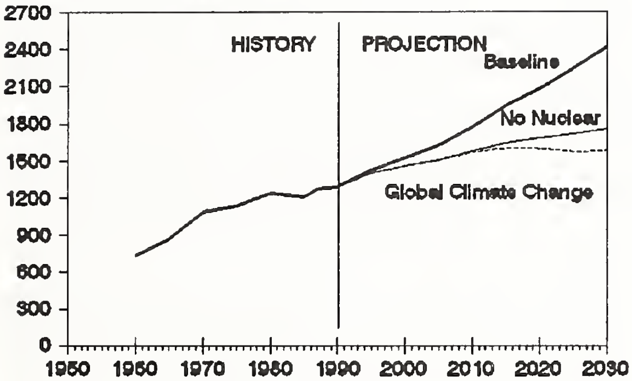


Figure 8. Sources of Oil Consumption
(Million Barrels Per Day)

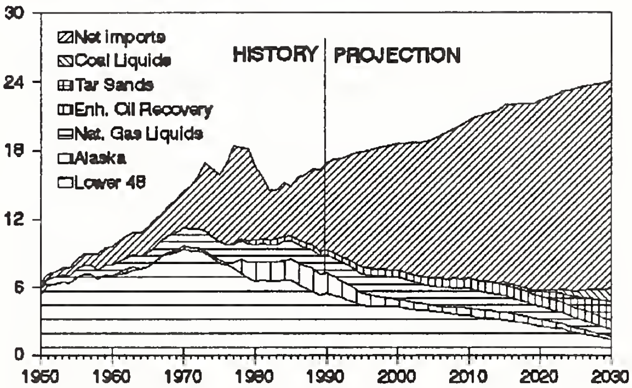
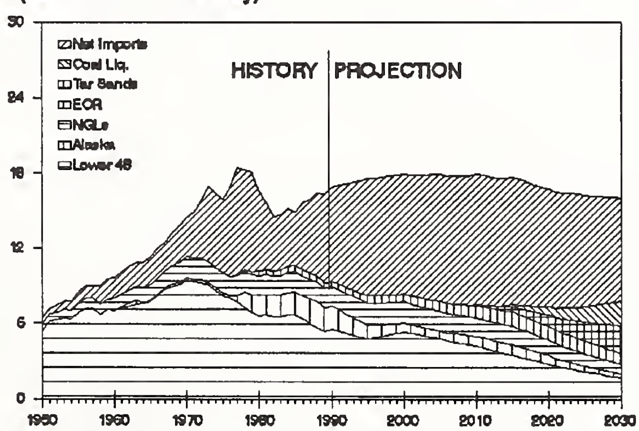


Figure 9. Oil Consumption: Vulnerability Case
(Million Barrels Per Day)



Methodological Issues in the Analysis of Macroeconomic Impacts of Energy Taxes

Ronald F. Earley & Kay A. Smith, Energy Information Administration

In the past year, the Energy Information Administration has been very involved in the assessment of the macroeconomic and energy market effects of alternative energy tax proposals. From a macroeconomic impact perspective, two methodological issues have evolved as being critical to the analysis. The first related to the role of accommodating fiscal and monetary policy and the second relates to the view about efficiency in the supply potential of the economy.

If the economic losses associated with the imposition of energy taxes are perceived to be large, accommodating policy options may be implemented. The mix of fiscal and monetary policies actually chosen will have differential effects with respect to inflation, employment, the Federal deficit, economic growth, and the composition of real GNP. These tradeoffs need to be subjected to explicit analytical scrutiny.

Also, alternative views about the resiliency of the economy can significantly effect the perception of the economic impacts associated with energy tax policy. Macroeconomic simulations were prepared using two assumptions regarding the effect of decreased energy consumption on the supply potential of the economy, defined as Potential GNP in the model.

The Modeling Process. The macroeconomic losses and Federal budget effects presented in this paper are based on simulations of the Data Resources, Inc. (DRI) U.S. Quarterly Model. For purposes of discussing the methodological issues, an incremental 50 cent Federal gasoline tax was levied on motor gasoline and diesel fuel. The tax was implemented using the gas tax lever in the DRI model. All subsequent energy price passthrough onto wholesale energy prices and resultant impacts on energy quantities were determined solely within the DRI model.

There was no attempt to bench energy prices and quantities to EIA energy model simulations. Therefore, the results that are presented below do not represent EIA's view of the impacts of the 50 gasoline tax. However, the pattern of the impacts, both in magnitude and the time-profile, are broadly consistent with other EIA work on taxes.[1]

The Role of Fiscal and Monetary Policy. During the course of the development of the National Energy Strategy, a debate ensued about the proper set of fiscal and monetary policies which should accompany any proposed energy policy. This is particularly important if there are large sums of collections from the tax. If the economic losses associated with the policy are perceived to be large, accommodating policy options may be implemented. The mix of policies actually chosen will have differential effects with respect to inflation, employment, the Federal deficit, economic growth, and the composition of real GNP.

Fiscal Policy. The simulations were carried out in two modes. In the Deficit Reduction mode, the gross collections from the imposition of the gasoline tax raise Federal indirect business tax revenues and these additional revenues are assumed to be used to reduce the deficit. For purposes of exposition, this simulation represents the central impact case. All of the alternative fiscal and monetary cases, as well as the discussion of the resiliency of the economy, are relative to the impacts in the deficit reduction simulation.

In the Deficit Neutrality mode, the gross revenues are not used to reduce the Federal deficit. The simulation explicitly targets the "full employment Federal deficit" to remain at baseline throughout the forecast period.[2] Two mechanisms for targeting deficit neutrality are investigated - a cut in the effective social security tax rate using a payroll tax cut and a reduction in the effective personal income tax rate. Changing the effective social security tax rate in the model implies changing both the employee and employer tax rates. The model was then targeted to achieve deficit neutrality by allowing the tax rates to change endogenously until neutrality was achieved.

Economic Impacts with Deficit Reduction. The imposition of the gasoline tax will have direct impacts on consumer spending and on the aggregate price level of the economy, with a complex time-profile of impacts on aggregate output. From the point of view of the consumer, two issues prevail. First, households will be faced with higher prices for gasoline and will adjust their amount of energy consumed. Nonetheless, nominal expenditures for gasoline will likely rise, taking a larger share of the family budget. Consumers can be expected to reduce their expenditures on other goods and services. Second higher energy prices effectively reduce real disposable income. Workers will attempt to bargain for higher wages. To the extent they are successful, this represents a cost increase in the production of all goods and services.

Energy services also represent a key intermediate input in the production of all goods and services.[3] After energy prices increase, substitution away from energy is limited and the prices of other inputs do not fall. This raises the production costs per unit of output for firms. Higher wage costs and spillover price effects on other variable costs further escalate production costs throughout the economy. This process places upward pressure on the nominal prices of all intermediate goods and final goods and services in the economy.

The effect on interest rates is subject to uncertainty, depending on the reaction of the monetary authorities. Lowering Federal government borrowing will place downward pressure on interest rates. However, if there is no change in the level of non-borrowed reserves, the higher aggregate price level effectively reduces the real supply of money. This will place significant upward pressure on interest rates.

On balance, the effect of the higher price level in the economy should dominate in the near-term forcing both nominal and real interest rates above baseline. The rise in interest rates discourages investment and consumer expenditures for interest rate sensitive components such as automobiles, housing and other durable purchases. In the long-term, as price inflation in the economy abates, the continuing reduction in Federal government borrowing serves to ultimately lower interest rates relative to the baseline.

The imposition of a gasoline tax collects an *ex-ante* level of revenues which is assumed to reduce the Federal deficit. However, with feedback effects on the economy considered, *ex-post* collections fall short of planned through the year 2001 (see the top of Figure 1). By 2002, with the economy nearly back to baseline activity, the reduction in the deficit matches collections. The rapid decline in the level of the Federal debt, coupled with a fall in interest rates and interest payments on the debt, actually serves to reduce the deficit by more than collections in the post-2002 period. The bottom of Figure 1 shows the GNP impacts.

Alternative Accommodating Fiscal Policies. The pattern of collections relative to reductions in the Federal deficit is important to consider in defining a deficit neutrality case. In

essence, "deficit neutrality" is not "fiscal neutrality". In order to achieve deficit neutrality, through 2001 less is returned to consumers or business than is actually collected. This is because the improvement in the deficit in the case discussed above was less than collections. In the post 2002 period, more is returned than collected. In this way, the federal deficit is maintained essentially at baseline levels.

As Figure 2 indicates, the return of collected tax revenues to consumers and business has the effect of moderating the GNP losses through the year 2000. The personal tax cut reduces the net present value of the real GNP loss to $170 billion, an improvement relative to the $199 billion loss in the deficit reduction case. The personal tax cut spurs consumption by increasing disposable income, but does nothing to lessen the underlying inflationary pressures set in place by the gasoline tax. The payroll tax, by initially cutting business costs, helps to control inflation. As a result, it is even more effective, reducing the GNP loss to a net present value of $61 billion.

However, if a longer time horizon is considered, the effectiveness of the deficit neutrality cases changes appreciably relative to the deficit reduction case. The personal income tax cut case actually yields higher GNP losses than the deficit reduction case when measured alternatively from 1990 through 2010 and through 2015. The aggregate price level remains above baseline. The improvement to consumption experienced in the first 10 years begins to wane. Worse, investment activity remains weak, particularly with interest rates remaining above baseline. The payroll tax cut significantly lessens the GNP impacts early in the forecast period, but by the year 2015 the economy has still not recovered to the baseline primarily because of higher sustained costs.

Monetary Policy. The theme is to devise monetary accommodation to achieve specific targets on real short-term interest rates. In the first case, reserves are changed to yield a return to baseline levels by 1992. The second case changes reserves such that real short-term interest rates are down by 50 basis points, compared to baseline, starting in 1992. This experiment maintains a certain parallelism with the deficit reduction/deficit neutrality cases by targeting a specific level of the monetary objective. These cases do not represent expectations about the degree of accommodation by the Federal Reserve Board. Rather, they represent some interesting sensitivity cases from which to view the relative economic impacts.

As shown in Figure 3, increasing reserves has the effect of moderating the GNP losses through the year 2000, but at the expense of moderately higher price level in the economy. Keeping short-term real interest rates at baseline reduces the net present value of the real GNP loss to $182 billion, an improvement of $17 billion compared to the baseline gasoline tax case. Pushing the real interest rate to 50 basis points below baseline yields a further improvement, with the net present value GNP loss dropping to $132 billion. The level of the CPI is higher in both cases.

As was found in the discussion of the fiscal policy cases, the view changes the longer the time-horizon. In the gasoline tax case (the deficit reduction case), interest rates fell significantly below baseline in the post-2000 period. Maintaining short-term interest rates at baseline essentially maintains interest rates at a higher level in the post-2000 period. As a result, the impact on GNP is actually worse than the underlying gasoline tax case. However, the 50 basis point case more closely tracks the underlying gasoline tax interest rate path. The loss in GNP never gets larger than the gasoline tax case, and actually continues to improve in the later years of the forecast period.

Summary. The comparison of alternatives within each arena (fiscal or monetary) is rather straightforward. Here, the use of a payroll tax cut combined with the gasoline tax seems to be preferable to the use of a personal income tax, particularly if the time horizon is only out to the year 2000. On the monetary side, the loss in GNP can be lessened. It just depends on how much additional inflation is acceptable and where is the appropriate target for interest rates.

In comparing monetary and fiscal policy, it still appears that the use of a payroll tax is most useful if accommodation of the gasoline tax is warranted. It lessens inflation while reducing the real GNP loss. Monetary policy will also reduce the GNP loss, but with higher inflation. If the magnitude of the monetary policy is large, as in the 50 basis point reduction case, the GNP loss can be lower than the payroll tax case considering the post-2000 period.

Efficiency in the Supply Potential of the Economy. The discussion above focused on the dynamic effects on the economy of changing energy prices and quantities alone, principally from a demand perspective. The ability of the economy to supply goods and services is equally important. **Potential GNP** is a measure of the output potential of the economy when all resources in the economy are fully employed — the unemployment rate is low, capacity utilization is high and there are no destabilizing events occurring in the economy. Potential GNP is forecast through an aggregate production function with labor, the capital stock and energy as key factors. Primary energy use goes down with increase in gasoline prices. What does this imply for the supply potential of the economy? Two views exist, with time a linking element.

- Reducing energy consumption alone does not necessarily improve the productive efficiency of the economy. The economy may be more efficient in its use of energy, but the lower level of energy use can adversely affect the productivity of other factors of production in the economy, thus lowering the potential output of the economy. This is particularly true in the near-term when the capital resource base is largely unaltered.
- However, in the longer-term higher energy costs would reduce the use of energy by shifting production toward less energy-intensive sectors, by replacing energy with labor and capital in specific production processes, and by encouraging energy conservation. There may be little or no sacrifice in the energy services rendered.

The assessment of the GNP impacts in the previous section essentially adhered to the view that a decline in primary energy use causes potential, as well as real, output of the economy to decline. The discussion which follows considers the premise that, in the long-term, there is no reduction in energy services, even though primary energy use is down. Essentially, this view states that the economy does not need to drop to a lower potential supply path simply because energy use is lower.

In the long run, it is the ability of the economy to increase supply that determines growth. The concept of potential GNP is crucial to describing the supply potential in the DRI model. Potential output in the economy is complex primarily because it is a theoretical concept, unlike other variables which can be observed and measured. To understand potential output, one has to compare it to its demand-side equivalent, real GNP.

First, the two concepts are different by definition. At most times, the economy does not fully employ all available resources in production — some people are unemployed and some machines lie idle. Real GNP is defined as the amount of goods and

services actually produced; potential GNP reflects the output of the economy using fully employed resources. This distinction makes potential output useful for analysis. The difference between potential and real output at any time can indicate the tightness of the aggregate economy. If real output is close to potential, shortages of labor and capital build up inflationary pressures. Wage costs and interest rates tend to increase. When resources are underutilized, the potential/real output differential indicates how much the economy can expand without risking inflationary tightness.

Second, real and potential output are calculated and projected differently. Real output is essentially a demand variable; potential output is a supply variable. Potential output calculations reflect how the national product is produced, not how it is used. The production function inputs are the analogues to the components of real GNP — consumption, investment, net export and government expenditures.

In theory, the procedure used to forecast potential output is fairly straightforward. One needs to forecast growth rates for each input to the production process and then determine how output will respond to the changes in the levels of inputs. However going from theory to practice, as usual, is a formidable task. Trying to aggregate diverse inputs into relatively homogenous categories is difficult. In addition, many inputs crucial to the production process may not be included adequately in the categories. Even after allowing for inaccuracies in representing diverse factors of production as aggregate inputs, questions remain regarding the manner in which efficiency changes are captured in the model. The economy's production potential is dependent on the quality of inputs as well as their quantity. However, existing input measures in the potential output equation may not reflect changes in input quality.

As a result, when trying to analyze economic impacts on the aggregate economy of various policy options, especially when the policies change a variable used in the formulation of potential output, treatment of input quality or efficiency becomes important. It is one thing to project potential output using smooth growth rate assumptions about the factors of production. It is another matter to examine the effects on the economy of large input price increases using analyses that suggest input efficiency changes will be large as well.

When these efficiency effects are taken into account the economy is essentially more resilient than when considering energy market price and quantity effects alone. When faced with the same percentage changes in energy prices and quantities, the dynamic effects on the economy are altered. First, labor productivity does not decrease by as much and labor costs do not increase compared to the unadjusted case. In turn, total prices do not increase as much leading to lower interest rates.

The quarterly model was solved using an alternative assumption regarding the effect of decreased energy consumption on the supply potential of the economy. Potential GNP was adjusted by the amount that energy consumption declined due to the gasoline tax (deficit reduction case). Thus, potential GNP would not fall as a first order response to the decline in energy consumption.

As a result, adjusting the supply potential of the economy has a positive effect on the level of the potential GNP impacts. As indicated in Figure 4, the impact curve of potential GNP is shifted upward, reflecting a reduction in losses early and even greater gains later. The price and quantity effects alone result in a net present value loss of $107 billion through 2015; if efficiency gains are incorporated, the net loss to the economy is $88 billion, an improvement of $19 billion.

Footnotes

[1] Energy Information Administration, *Studies of Energy Taxes,* (SR/EMEU/91-02)(April 1991).

[2] The Economic Policy Council on Energy Taxes advocates the deficit neutral approach for the assessment of energy tax policy options.

[3] Although gasoline plays a smaller role as an intermediate input to the production process, other taxes such as a Btu or carbon tax would influence prices through this channel.

Table 1. ECONOMIC IMPACTS

	Net Present Value of Real GNP Loss, Billion 1982 Dollars		
	1990-2000	1990-2010	1990-2015
A. FISCAL POLICY			
Deficit Reduction	-199	-154	-107
Personal Tax Cut	-170	-241	-240
Payroll Tax Cut	-61	-132	-148
B. MONETARY POLICY			
At Baseline	-182	-159	-115
Down 50 Basis Points	-132	-76	-12
C. EFFICIENCY GAINS			
Higher Potential GNP	-188	-132	-88

Figure 1

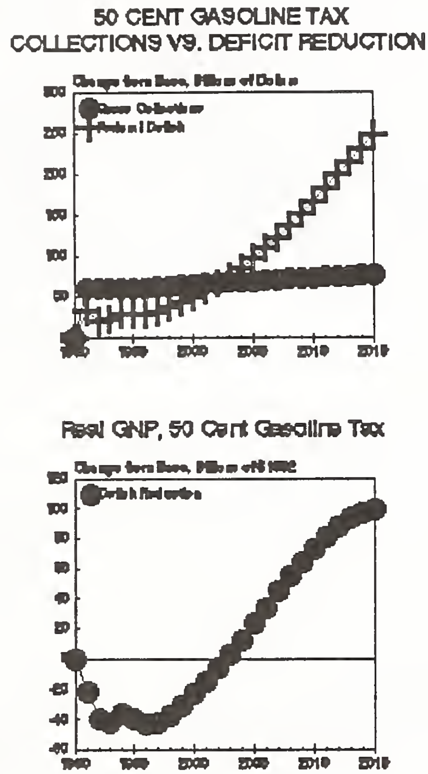
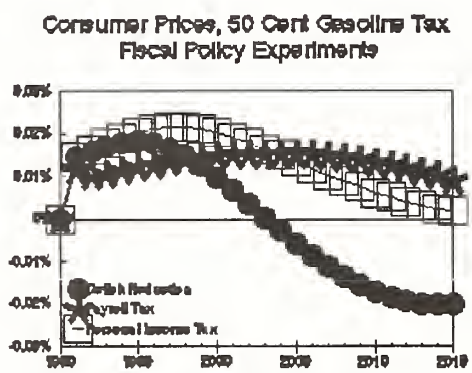
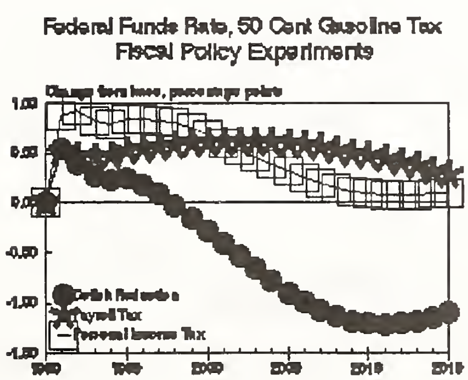


Figure 2

FISCAL POLICY EXPERIMENTS

DEFICIT REDUCTION
PAYROLL TAX CUT
PERSONAL INCOME TAX CUT





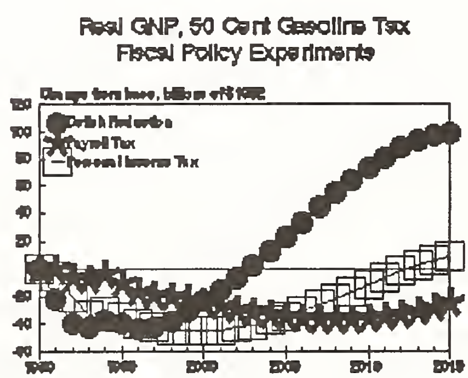
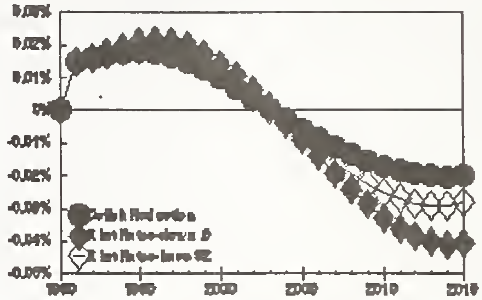
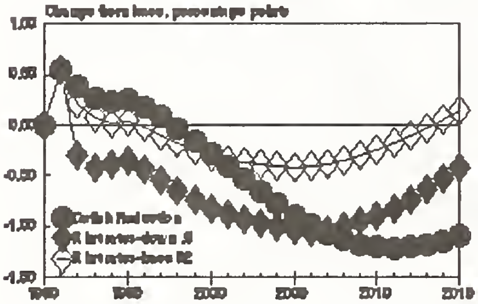


Figure 3
MONETARY POLICY EXPERIMENTS
REAL INTEREST RATES AT BASELINE
REAL INTEREST RATES DOWN
50 BASIS POINTS

Consumer Prices, 50 Cent Gasoline Tax
Monetary Policy Experiments



Federal Funds Rate, 50 Cent Gasoline Tax
Monetary Policy Experiments



Real GNP, 50 Cent Gasoline Tax
Monetary Policy Experiments

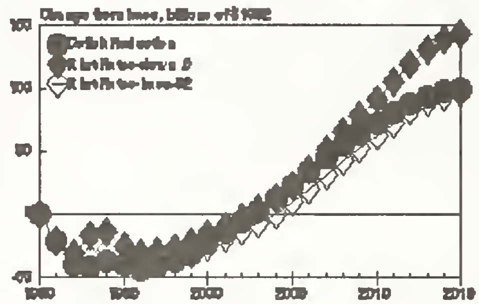
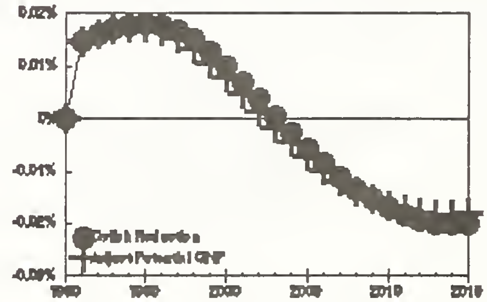
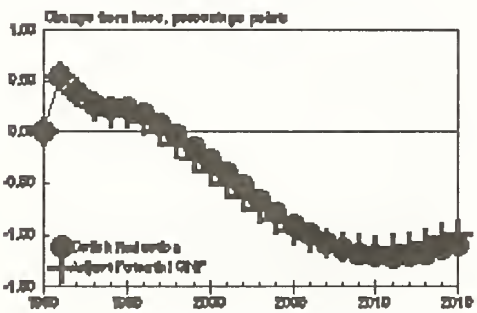


Figure 4
ENERGY EFFICIENCY EXPERIMENT
POTENTIAL GNP ADJUSTED

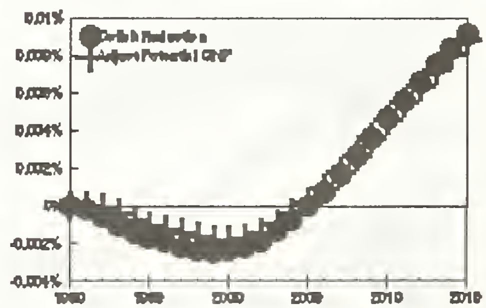
Consumer Prices, 50 Cent Gasoline Tax
Energy Efficiency Experiment



Federal Funds Rate, 50 Cent Gasoline Tax
Energy Efficiency Experiment



Real Potential GNP, 50 Cent Gasoline Tax
Energy Efficiency Experiment



The Development of the National Energy Modeling System

C. William Skinner, Energy Information Administration

Developing and using large-scale energy forecasting models have been major responsibilities of the Energy Information Administration (EIA) and its predecessor organizations since the Project Independence Evaluation System (PIES) was constructed in 1974. Over the years, the forecasting models have been used to prepare annual forecasts for the period 10 - 20 years in the future (the period we now call the midterm), which are published in the Annual Energy Outlook, and for innumerable studies of the effect of proposed energy policies or other uncertainties on key energy variables.

The forecasting system currently in use, the Intermediate Futures Forecasting System (IFFS) became operational over ten years ago, when energy markets, and the hardware and software available to model them, were quite different than they are today. The various models which are components of IFFS have been revised, updated and improved, but it became apparent several years ago that major revisions and additions would be needed to address the energy policy issues becoming increasingly important. EIA's involvement in the National Energy Strategy analysis disclosed additional requirements for long-term forecasting (20 - 40 years in the future) and for environmental and renewable energy capabilities well beyond those available in IFFS. In addition, the models needed to be designed and programmed in such a way that they could be easily understood, easily modified, and provide rapid response for exploring alternative policies.

In the Fall of 1990 EIA initiated a project to develop a new National Energy Modeling System (NEMS) which will update its forecasting capability and extend and enhance it to meet the many new and anticipated requirements. The stated objective of the NEMS is:

To illustrate the energy, economic, environmental and energy security consequences on the U.S. of various energy policies and assumptions by providing forecasts of alternative energy futures in the near-term, midterm, and long-term periods, using a unified system of models.

To be effective NEMS must play several roles. It must provide the base-line forecasts of the U.S. energy future which are published in the Annual Energy Outlook. It must be sufficiently detailed to simulate the regional and sectoral effects of a wide array of energy policies which are expected to be of interest in the next decade — policies such as energy taxes, financial incentives, regulations of various kinds — as well as a range of assumptions concerning economic growth, demographics and the impact of new technologies. Yet many of the uses of the modeling system do not require such extensive detail, but do require the ability to explore a variety of options rapidly and easily. A major challenge in the NEMS development is to design a system which can satisfy the competing requirements for extensive detail and rapid operation.

The development of the future NEMS is in the early stages and many details of the design have yet to be resolved. Still, it is possible at this stage to give a reasonably clear picture of the overall system structure and the ways in which the system will be used to produce forecasts.

Overall System Architecture. The NEMS design employs a highly modular, highly flexible approach which will allow the system to be configured to suit the needs of a particular analysis. Each individual module of the system can be used in its fully detailed form or replaced with a simplified, "plug-compatible" version for simplicity and rapid operation. By assembling an appropriate collection of modules, the system can be tailored to provide detail only where it is needed.

In some cases the full regional and sectoral detail of the system will be needed and can be used. In others, where such detail is not required, it can be collapsed into a more aggregated treatment, resulting in a faster, simpler model configuration. With appropriate guidance from the user interface portion of the system, the user can tailor the system configuration to fit a particular issue by providing greater or lesser detail in selected areas. It has often been said by professional energy analysts and modelers that the best results are obtained when a model is designed to study a particular type of issue. Yet NEMS in its entirety must be applicable to a wide range of proposals and policies, and to all widely-used forms of energy. By using a building-block approach, with great flexibility in the ways the blocks can be put together, it is hoped that both comprehensive coverage and precise fit to a given issue can be obtained. To accomplish this flexibility of operation the system must be structured to assure maximum separability of the modules and the interfaces must be cleanly and carefully defined.

The NEMS is an integrating system, which balances supply and demand for all energy products concurrently, allowing fuel substitution and feedback effects among energy markets. The major, top-level modules of the NEMS reflect the dominant energy markets and consuming sectors in the U.S., as shown in Figure 1, together with a macroeconomic module for interaction with the U.S. economy and an international module for interaction with world energy markets. An integrating model is required to control and balance the results of these major modules.

Each module is constructed of sub-modules, which in turn may have sub-modules, down to a level where the sub-module operation is simple and apparent. At this stage of the design process, only the top-level and, in some cases, the second-level modules have been defined. Each module at each level can be used in its most detailed version or replaced by a more aggregated version with corresponding inputs and outputs. By instructing the system to include more or fewer detailed modules for a given analysis, a user can tailor the NEMS to fit the analysis.

To focus on a particular area of interest, each top-level module will be designed to operate with the simplest possible versions of the other modules or even totally independently, with data and assumptions replacing the inputs normally received from other modules. In this way, a study which focuses on coal, for example, might employ the detailed versions of the coal and electricity modules and aggregate versions of all the rest. By specifying the use of detailed representations only where needed for a given issue, NEMS can achieve significant improvements in both responsiveness and transparency. Responsiveness is measured in part by how rapidly a given simulation runs on the computer, but a better measure is the time required to set up, run and interpret the results of a given simulation. Transparency reflects how easy it is for a user to understand why the model behaves as it does — which assumptions or inputs to the system have a significant impact on the results.

The system framework is designed to facilitate further growth and development over time, as well as to allow for steady evolution to meet changing energy markets and changing analytical demands. If an important application arises for which NEMS is not

a suitable tool, it will often be possible to modify or expand one or two small modules, without impacting the rest of the system, and have NEMS become useful for that application.

User Interface. NEMS is aimed at experienced energy analysts who are familiar with the subject matter and the goals of their analysis, but who may not be model builders or be intimately familiar with the details of the NEMS construction. The user interface portion of NEMS will give such a user easy access, guided if desired, to the assumptions and inputs he may wish to modify for a particular model run. Emphasis will be placed on allowing the user to specify the conditions for a given model run, perhaps in great detail, but not requiring the user to specify anything. (In this case the standard assumptions would apply.) Since the user interface will make use of modern interactive computing techniques, specifying the conditions for even the most detailed simulation can be rapid and efficient.

The time to actually execute the modules assembled for a particular simulation is a function of many things, including the computing hardware available and the skill with which the modeling algorithms have been implemented. The user can control the execution time to a large extent, however, through his decisions at set-up time on the degree of detail to be employed in the various energy source and consumption modules and on the number of years in the forecast period. Execution times in minutes are anticipated for the least detailed configurations; perhaps in hours for the most detailed.

The form and format of the output from NEMS will greatly aid the interpretation of simulation results. The interactive user interface will provide the tools to examine selected outputs in graphical or tabular form, to compare the results with those of previous runs using different assumptions and, hopefully, to get a sense of the uncertainty associated with particular outputs.

Needless to say, the responsiveness of a forecasting system is more than just how quickly a model run can be set-up, executed and interpreted. A very simple model can provide results very quickly, but those results may be useless because they are not sufficiently sensitive to the assumptions being examined—"quicker is better" if, and only if, detail is provided which is adequate for the issues being considered.

Central Data Base. To help assure separability of the modules and to provide for overall system management, a centralized data repository will be used as the principal means of communication among the modules. The data base will contain three types of data:

(1) Fundamental input data and assumptions for use in common by all the modules

(2) Outputs from the modules, which may be used as input by another module or used for final forecast reports

(3) Metadata — data about the data — which allow an analyst to track, in descriptive terms, the assumptions and conditions which lead to a particular forecast

In the EIA environment, where NEMS will likely be in use by several analysts at once, it is expected that this data base will be maintained on a large mainframe computer, accessible to the personal computers through a Local Area Network. For users in other environments who may not have access to a mainframe or other suitable file server, it may be preferable to maintain the data base on optical disks or similar bulk storage devices attached to the personal computer. Many questions such as this, related to choices of computer hardware and software, remain to be settled.

Conclusion. The Energy Information Administration has undertaken the development of a new National Energy Modeling System to update, extend and enhance the integrated energy forecasting capabilities which have been available to the Department of Energy. To evaluate the impact of many energy policies of potential interest, detailed regional and sectoral representations are required, leading to large and complex models. For initial exploration of policy options and for general sensitivity analyses, transparency to the analysts and rapid turnaround are essential. The new NEMS aims to satisfy these competing requirements by employing a highly modular, highly flexible design which allows various configurations of detailed and simplified modules to be assembled, depending upon the needs of a particular analysis. The overall system design for the NEMS will be completed soon, with detailed design of the individual modules and implementation proceeding over the following two years.

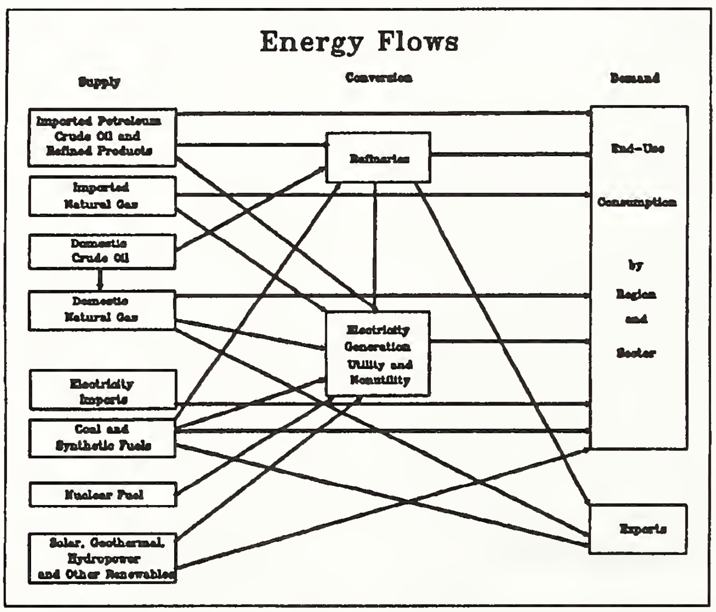


Figure 1.

The Availability, Utility and Comparability of Federal Data to BHPr

Sandra Gamliel, Bureau of Health Professions

The Bureau of Health Professions (BHPr), a component of the Health Resources and Services Administration (HRSA), collects and analyzes data on health personnel and health professionals in particular. BHPr staff assess the nation's health sector work force and projects future supply and requirements using a variety of methods. The Bureau routinely prepares or supports studies and reports on the supply, distribution, and utilization of practicing health professionals and health professions' students and institutions.

In order to carry out its function of monitoring the current and future supply of and requirements for health personnel, BHPr has tapped data from a number of sources. Current sources of data include studies and surveys conducted by in-house staff, such as the Sample Survey of Registered Nurses, and professional associations, most notably the American Medical Association (AMA), the American Hospital Association (AHA), American Dental Association (ADA), and the National League of Nursing (NLN).

In recent years, major changes in the health care sector have affected the supply and utilization of health care providers. These changes have occurred during a period of time when resource constraints have limited the amount of primary data collection which BHPr may conduct. As a result, the Bureau of Health Professions has begun to examine the utility of using other data sources in an effort to continue its mission of monitoring health personnel.

While data and studies from associations and from selected Federal agencies are used extensively, other federal data sources are perhaps underutilized. In some instances constraints imposed by sampling strategies, methods of data collection, data quality, and data availability may limit the utility of these sources to BHPr. In other cases however, federal data sources currently not being used may enhance BHPr capabilities. BHPr contracted with Kunitz and Associates to assess the utility and relevance of existing data sources in other federal agencies to BHPr programs and objectives. This presentation describes their findings. As noted earlier, BHPr, currently deals extensively with many professional associations and is on the whole cognizant of non-government sources of data. Therefore, this study was limited to only federal data sets; specifically those pertaining to the supply and demand of health personnel.

SUPPLY. The future supply of any specific type of worker is dependent upon three factors: the new entrants into the field; the aggregate number, personal characteristics, and geographic distribution of persons currently practicing, and the number of persons leaving the occupation due to death, retirement, or other reasons. It is of particular interest to examine these factors in relation to health care settings, and the specific geographic locations of providers and facilities when possible.

New Entrants. Data on new entrants into the health care field is discipline specific and therefore, varies by occupation. Much of the data currently used by BHPr comes from professional associations such as the AMA, the AAMC (American Association of Medical Schools, NLN (the National League of Nurses) among others.

One Federal data source on new entrants is the IPEDS or Integrated Post Secondary Education Data System available from the National Center for Education Statistics. Data on the number of schools with programs in specific fields, and the number of students and graduates by field of study is collected in this survey. While the data may not be as detailed for some occupations such as physicians, as that provided by the professional association, this data may be the only information available in some of the allied health fields.

Aggregate Supply and Separations There are a number of Federal databases with contents relating to the current supply, demographic characteristics and/or distribution of the labor force and health personnel. Chief among these sources are the Bureau of Labor Statistics (BLS), National Center for Health Statistics (NCHS), The Bureau of the Census.

The Bureau of Labor Statistics collects and compiles data on facilities and personnel. BLS ES-202 Earnings and Wages) and 790 data (Employment and Earnings) provide information on industry trends down to the four and three digit SIC levels respectively. Data on reporting units and employment is available from these sources. In addition, BLS collects information on staffing patterns by 3-digit SIC industry groupings in the OES (Occupational Employment Statistics) Program. Furthermore information on detailed occupations, demographic characteristics and separation rates is available from the Current Population Survey. BLS also conducts an industry wage survey which periodically surveys wages paid to selected health personnel in selected health industries.

The National Center for Health Statistics collects data relative to both supply and requirements for health personnel. The discussion of data relative to the demand and requirements for health personnel will be deferred until later. There are two major sources of supply data available from NCHS: The National Ambulatory Medical Care Survey (NAMCS) provides data on the availability of physician providers in the office-based setting, while the National Nursing Home Survey (NNHS) contains valuable data on nursing personnel in the nursing home setting.

In addition to the occupational and characteristics data available from the decennial Census the Bureau of the Census also provides data on the location of facilities, number of employees and annual payroll. The Census of Service Industries (CSI) and County Business Patterns (CBP) provide data by industry SIC codes.

REQUIREMENTS. The analysis and forecasting of demand or requirements for health personnel depends extensively on past trends in the consumption or utilization of health services. Data on the utilization of health care services by sector is available from a number of sources.

The Agency for Health Care Policy Research (AHCPR) is responsible for two studies relevant to describing and analyzing utilization patterns and forecasting health personnel needs. The Hospital Cost and Utilization Project (HCUP) survey examines utilization in the hospital sector and the National Medical Expenditure Survey provides national estimates for access to care as well as utilization patterns.

The Alcohol Drug Abuse and Mental Health Administration (ADAMHA) conducts the Inventory of Mental Health Organizations and General Hospital Mental Health Services which provides information on inpatient and outpatient mental health treatment facilities. Data on substance abuse treatment centers

which is not directly relevant to BHPr is also available through ADAMHA's National Drug and Alcoholism Treatment Unit Survey.

NCHS conducts four separate surveys relevant to requirements analysis and forecasting: The National Nursing Home Survey, The National Ambulatory Medical Care Survey, The National Hospital Discharge Survey and the National Health Interview Survey. The surveys provide estimates of utilization among the various health care service sectors along with patient profiles.

A wealth of utilization data for those persons 65 years of age and over is available from the Health Care Financing Administration (HCFA). Detailed information on patient demographics, residence, diagnosis, procedures, length of stay and discharge disposition is available from HCFA's MEDPAR file.

CONCLUSIONS AND RECOMMENDATIONS. The surveys and data bases reviewed can provide a great deal of information relevant to BHPrs needs. Data bases that deal with the supply of and requirements for health manpower are clearly available in the public sector and may be used to augment the data currently used and fill some of the data gaps that have hindered BHPr in carrying out its mission of monitoring and forecasting the supply of health care professionals.

It is recognized that these data may not be fully comparable to those used by BHPr and the data are not collected to address the specific issues of interest to BHPr. Therefore, attention to the following aspects of each data base reviewed is essential in gaining an understanding of its applicability and utility:

- Purpose of the data collection
- Data collection methodology
- Source of data
- Data element definition
- Availability of data base

Once such an examination has been completed and the utility of the database assessed, cooperation between BHPr and the federal agency collecting the data is essential in order to facilitate a clear understanding of the data's limits and allow for easy access to the data.

Expanding the Health Personnel Components of Secondary Data Sources

Edward S. Sekscenski, Bureau of Health Professions

In response to the need for comprehensive information on the supply of and future requirements for health personnel in the United States, the Bureau of Health Professions relies heavily on the utilization of secondary data sources, especially those gained from several survey efforts of other Federal and non- Federal health services research and statistical agencies. Much of the information the Bureau is required to submit in its biennial Report to Congress and the President can be obtained from ongoing data collection programs. Some of these, for example, are discussed in the other two presentations being made here today. Ongoing data programs used by BHPr in its modeling and forecasting activities include those of the Bureau of Labor Statistics, the Bureau of the Census, the American Hospital Association, the American Medical Association, and various other programs of health professions associations and other organizations. Further information, however, is often necessary for answering questions requiring more specificity in regards to the type and number of health personnel needed to address particular health care problems facing the Nation.

Recently, staff of the Office of Data Analysis and Management of the BHPr have participated in the development of several important survey efforts focusing on provision of health services to special population groups and on issues of major public health concerns. As a result of these efforts, new national data are or will soon become available on health personnel providing services to substance abusers in alcohol and drug treatment programs as well as to persons receiving formal health care services in their homes. This paper discusses the development of health personnel data collection within specific secondary data sources addressing these two areas. It also will describe, in brief, use of some of these collected data within some of the Bureau's programs and reports.

HEALTH PERSONNEL PROVIDING TREATMENT SERVICES FOR ALCOHOL AND OTHER SUBSTANCE ABUSE

Despite some evidence of a decline in the overall use of illicit drugs in recent years, both alcohol and drug abuse remain major public health problems in the United States, impacting profoundly on the health care needs of the Nation. Alcohol and other substance abuse are factors in about half of all U.S. traffic accidents, up to two-thirds of deaths from falls and fires, and are implicated in a large proportion of criminal activity resulting in bodily injury and death. Together with related mental and physical co-morbidities substance abuse contributes heavily to increased disability, chronic illness, and premature death in the United States. One form of substance abuse, that which entails intravenous drug use (IVDU), is especially affecting the public health of the U.S. population as an increasingly significant risk factor in the spread of the human immunodeficiency virus that causes AIDS. Clearly the consequences of alcohol and drug abuse are major health problems for which the need for trained and qualified health personnel will continue throughout this decade.

The Bureau of Health Professions has reported on the health personnel needs of the Nation's alcohol and drug abuse treatment system in previous reports. However, no current national data on the supply of these personnel was available to BHPr at the time of preparation of its Seventh biennial report[1]. Because of this lack of information upon which to review current resource use and project future health personnel requirements for addressing the problems of substance abuse, staff of the Bureau of Health Professions have recently pursued and obtained input into two survey efforts sponsored by the PHS's Alcohol, Drug Abuse and Mental Health Administration (ADAMHA).

The ***National Drug and Alcoholism Treatment Unit Survey*** (NDATUS) collects data on the facilities, clients, capacity, and funding sources of all public and private substance abuse treatment units in the United States. The NDATUS has been conducted periodically by ADAMHA's National Institute on Drug Abuse (NIDA) and National Institute on Alcoholism and Alcohol Abuse (NIAAA) using a number of survey instruments since the mid-1970s. Data are available from these surveys for various years 1974-1990.[2]

Upon learning that the survey instrument for 1989 NDATUS was to be revised and that the survey would be fielded again in both 1990 and 1991, the Bureau of Health Professions asked the ADAMHA if we could participate in expanding the information collected on the staffing patterns of treatment units. We were successful in this endeavor only to the extent of obtaining further delineation of psychiatric specialty within the medical profession and to record the total hours worked by all part-time personnel. The latter, however, allows for a better estimation of paid-staff FTEs employed in drug treatment units. The former provide the BHPr with additional information for forecasting physician requirements for delivery of these health care services.[3] Health personnel collected in the 1990 and 1991 NDATUS include the occupational categories of psychiatrists, other physicians, registered nurses, other medical, psychologists, social workers, credentialed counselors, and all other direct care providers. Preliminary data from the 1990 NDATUS were made available from ADAMHA in August 1991.

The slight expansion of the staffing matrix represented a compromise between BHPr and ADAMHA in regards to the additional occupational detail that could be included in this survey. The NDATUS is an attempt at a census accounting of drug and alcohol treatment units, and detailed data collection is therefore limited in it. We appreciate the additional information collected in the 1990 and 1991 surveys relating to staffing of these facilities nationwide. The data needs of the BHPr models and programs, however, require more detailed information with respect to both the occupational mix of health personnel providing services in the Nation's substance abuse treatment system, the organizational structure of the facilities providing this treatment, and the characteristics of patients/clients receiving such treatment.

Partially through our work with ADAMHA on the revision of the NDATUS form, ODAM was asked to participate in the development of a more detailed survey of drug and alcohol and drug-only treatment units that utilizes the NDATUS as its sampling frame.

The ***Drug Services Research Survey*** (DSRS) is a national sample survey of drug treatment facilities and their clients co-sponsored by NIDA and the Office of National Drug Control Policy (ONDCP) (in coordination with a number of state and national associations of substance abuse treatment providers). The DSRS, which was fielded in the Spring of 1990, provides detailed facility-level information from telephone interviews with 1,000 of the approximately 7,000 drug-only and drug and alcohol treatment units identified in the 1987 NDATUS (supplemented

with information from the 1989 NDATUS) and client-level data from on-site abstracts of approximately 20 client records in each of a subsample of 120 of these facilities.

Health personnel staffing information was collected in the DSRS on the facility component of the survey as well as in a followup face-to-face survey with administrators. ODAM has worked very closely with both NIDA and Brandeis University, a subcontractor on the study, in developing the staffing-patterns questions included in the survey instrument for the DSRS. ODAM is continuing to provide technical assistance in analyzing the data that has been collected.

The DSRS survey provides information on occupational detail and employment status of health personnel providing services in substance abuse treatment facilities according to their payroll, contracted services, or volunteer status, and expands upon that collected in the NDATUS. The survey also allows for analyses of staffing patterns of treatment units in relation to detailed facility and client- specific characteristics, representing an expansion of the research possibilities not available in prior national surveys. The following is a brief description of information obtained from the DSRS, some of which will also be presented in the upcoming Eighth Report to the President and Congress:

> Beyond providing health care to persons suffering from poor health due to the physical effects of alcohol and substance abuse, a large number of health professionals are involved directly in providing substance abuse treatment. This treatment is provided through various modality types, including hospital inpatient and outpatient units, residential care programs, outpatient drug free maintenance programs, alcohol or drug-abuse only programs, multiple-dependency programs, and inpatient detoxification programs. According to the Drug Services Research Survey, a total of about 68,700 health personnel were employed directly by these treatment units in 1990. About seven of every ten of these employees worked full time. Another 13,500 health personnel also worked for these alcohol and drug treatment units on a contractual basis. The type and number of health personnel working in alcohol and drug treatment units are shown in Table 2.

HEALTH PERSONNEL PROVIDING HOME HEALTH CARE SERVICES

Another area of particular interest in health services research is that related to the growing field of home health services delivery, especially to the elderly and the chronically ill. Continuing a long-standing relationship between the two PHS agencies, staff of BHPr have been working with staff of the National Center for Health Statistics (NCHS) in their development of survey instruments that are or will collect data on health personnel employed in home health care agencies. In particular, various BHPr staff have provided technical assistance with respect to adding specific questions to the NCHS-sponsored Health Provider Inventory as well as the pretest of the Home Health Care Survey, a long-term-care component of the developing National Health Care Survey.

The HPI is an attempt at a national census of health care providers that may be utilized by NCHS in the sample frame development for many of their other provider surveys. The data items collected are limited by and large to what is necessary for stratification purposes and other sample design issues. We were very pleased, therefore, to have input into this important survey instrument. Obtainment of even limited information on the health personnel staffing of various types of health care providers in the HPI allows for the potential to utilize personnel as one of the stratifying criteria in future survey efforts.

Of course, data can also be utilized directly from the Health Provider Inventory in an analytical framework. Staffing level information from the long-term-care component of the HPI, for example, will likely provide BHPr with information necessary to estimate the type and number of personnel employed by home health agencies. Disaggregating the mix of these health personnel according to various home health agency characteristic, for example by size, geographic location, and licensure status, will enable BHPr to get a handle on the possible impact of health personnel demand changes likely to result from projected growth in these health care providers in the coming years.

The last survey effort I wish to discuss with you today is the pre-test to the NCHS Home Health Care Survey. This pretest, scheduled to be fielded this fall by the Bureau of the Census for the National Center for Health Statistics will collect data from a sample of home health service agencies to be drawn from the Health Provider Inventory. Staff from both the Division of Nursing and the Office of Data Analysis and Management, of BHPr, assisted NCHS in the development of a small number of questions on this survey. These questions will obtain staffing level information as well as information on visits made by health personnel to persons receiving services from the home health agencies surveyed. This information will be utilized to develop the final survey instrument for a more comprehensive Home Health Care Survey, hopefully to be fielded in the next year. From the latter survey, BHPr anticipates receipt of data that will enable us to conduct more detailed analyses of such issues as the type of care provided by home health care agencies utilizing differing allocations of staff resources.

Through our participation in the development of these and other secondary data sources BHPr seeks to obtain information needed to address issues impacting on the supply of and requirements for health personnel to meet the changing health care needs of the U.S. population.

Footnotes

1 U.S. Dept. of Health and Human Services, ***Seventh Report to the President and Congress on the Status of Health Personnel***, Public Health Service, Bureau of Health Professions, March 1990.

2 Data from the NDATUS are available for 1974, 1979, 1980, 1982, 1984, 1987, 1989, and 1990.

3 U.S. Dept. of Health and Human Services, ***Highlights From the 1990 National Drug and Alcoholism Treatment Unit Survey (NDATUS)***, ADAMHA, National Institute on Drug Abuse and National Institute on Alcohol Abuse and Alcoholism, July 1991.

Table 2. Preliminary estimates of the number and type of health personnel providing services to persons in drug and alcohol or drug-only treatment units in the United States, March 30, 1990.

Health Occupation	Payroll Employees		Contract	Total
	Full-time	Part-time	Personnel	Personnel
Registered Nurse	9,344	4,002	871	14,217
Licensed Practical Nurse	4,740	1,662	475	6,897
Non-psychiatric Physician	1,043	1,888	2,835	5,766
Psychiatrist	933	1,994	2,497	5,424
Psychologist	2,565	1,657	2,486	6,708
Social Worker	5,164	1,603	1,545	8,312
Family Therapist	2,961	987	627	4,575
Vocational Rehabilitation Specialist	777	240	242	1,259
Other Degreed Counselor	19,549	3,129	1,621	24,299
All other Health Personnel	3,181	1,272	342	4,795
Total Health Personnel	50,257	18,45	13,541	82,252

Source: Drug Services Research Survey, presented in provisional report prepared for NIDA by the Bigel Institute for Health Policy, Brandeis University April 1991.

Using and Building on Association Data Bases in Federal Sector Health Personnel Analysis

Herbert Traxler, Ph.D., Bureau of Health Professions

The author gratefully acknowledges the comments and input by Gloria Bronstein and Jim Cultice, especially to the Dental and Physician sections, respectively. Any remaining errors or omissions are the author's.

Introduction. The Bureau of Health Professions—especially our unit, the Office of Data Analysis and Management (ODAM)—depends primarily on other sources for basic data on health professionals. These sources include both the Federal Government and professional associations. This paper covers databases from associations; another paper in this session discusses databases from Federal Government sources.

The presentation will deal with a few selected associations' data bases, and some of our contractual and informal relationships enhancing our use of these data. It is illustrative rather than comprehensive, does not cover all their data nor all of our use thereof, or of data of other organizations or associations; nor do I intend to cover our analysis of these or of other major organizations' or professions' data. Not covered is nursing, for example, the only occupation where the Bureau has undertaken primary data collection in the 1980s (in 1983 for licensed practical nurses, LPNs, and for registered nurses in its 1980, 1984 and 1988 Nurse Sample Surveys).

The Importance of personal contacts through site visits. The data is obtained from these associations in part through contracts, in part through purchase of their publications, or through official or informal channels. These were established and are maintained in regular (mutual) telephone and written communication, and mutual visits by lead analytical staff of the Bureau and the associations. An example of these is a recent site visit on April 4-5, 1991 to the home base of four major professional associations in Chicago, in which I was accompanied by Mr. Jim Cultice of my staff. During those two days we met first with the American Osteopathic Association (AOA) to discuss potential for our contractual use of their 1988 census data, then with several components of the AMA to discuss their young physician survey, especially as a source for minority data, discrepancies in FMG numbers and some other issues related to our versus AMA analyses and forecasts of physicians, the current professional activities survey, conveying our interest in obtaining minority data - especially by practice location - and the status of our current contract on "Basic Data on Physician Characteristics." With the AHA's Hospital Data Center staff we discussed analytical issues surrounding our special analyses of and reports on ***Trends in Hospital Personnel***, and their special surveys of human resources, both of which are valuable data sources especially for the data-poor allied and associated health professions. Finally, association and other data for our Econometric Model of the Dental Sector (EMODS - see below) and the possibility to have more race/ethnic specificity in the ADA's survey of dental practice were explored with ADA analysts. It is hardly possible to overstate the impact and cross-fertilizing yield from such site visits, and the continuing collegial relationship between federal and key private analysts such as those of the four mentioned professional associations.

In the following few pages I shall highlight a few of these association sources which are the primary custodians of analytical data for several major professions under the purview of the Bureau's analysis, and briefly illustrate our use of these data.

Dentistry.

The Bureau of Health Professions (BHPr), HRSA, USPHS utilizes two computer models in its analysis of dental personnel. The Econometric Model of the Dental Sector (EMODS) is a very complex econometric model which forecasts a number of dental care components, and ultimately is used to estimate future dental personnel requirements. This model (EMODS) has been recently updated, and was discussed in more detail in the previous two Federal Forecasters Conferences (FFC-89 and FFC-90 - see the ***Proceedings for FFC-90***).[1] EMODS in turn uses as one of its many inputs the projected number of dentists by age, which is produced by another Bureau model, the BHPr dental manpower supply model. These two models have at times been mistakenly referred to as one model in the literature, leading to confusion and erroneous interpretations of their respective premises and findings.[2]

Main Features of BHPr's Dental Models. The computer-based dental manpower supply model estimates the number of dentists entering and leaving the profession over the projection period. The baseline data are from the American Association of Dental Schools' (AADS) applicant and enrollment figures and survey of dental school seniors. Different assumptions about future enrollments produce several projection levels for the supply of dentists. In addition, uncertainties about future economic conditions, economic challenges to dental schools themselves, future applicant pools, and changes in dental disease patterns will indirectly affect the supply of dentists. However, changes in these conditions are not currently part of the supply model.

EMODS is an economic equilibrium model, whose major components are the demand for and supply of dental services. Interaction between these components causes the cost of dental care to move to a level where dental services supplied equal those demanded. The model results in calculated future values for expenditures for dental services, cost and utilization of dental services, dental income, price of equipment and supplies, and employment levels of dental auxiliaries. EMODS forecasts are produced within the model's framework of interactive mathematical equations under specific assumptions about future underlying conditions and governmental policies that affect these variables. Detailed descriptions of the model, including flow diagrams illustrating these interrelated factors and variables, are contained in the literature.

Data from Associations. The following data for EMODS variables are from associations (i.e, non- Government) sources:

Dentist Supply. As already noted, dentists by age for our EMODS model are obtained from the Bureau's dental manpower supply model. The American Dental Association (ADA) supplies the historic figures by age every three years. The Bureau interpolates the ADA's total figures for intermediate years, then distributes the totals by age for intermediate years. The supply model then projects dentists by age, subtracting an estimated number who will probably retire or die (using age-specific attrition rates) and adding new graduates which are projected from first-year enroll-

ment and graduate figures obtained from the American Association of Dental Schools (AADS). All other data inputs are handled directly by our EMODS econometric model.

Dental Hygienist Supply. Updated figures on dental hygienist first-year enrollees and graduates are from the ADA's "Annual Report on Dental Auxiliary Education 1988-1989". These updates add to the historic database on supply of hygienists.

Dental Wages and Office-Related Expenses. The ADA's Survey of Dental Practice provides the data on average annual wages of hygienists, dental assistants, and clerical support. This is also the source for figures on cost per dental chair, as well as supply costs, lab charges, drug costs and number of services which determine the "unit cost".

Dental Productivity. The ADA has also been very helpful this year in our investigation of EMODS' assumption for neutral technological change, and its corollary factor, dental productivity; this supplemented background information we received from the Bureau of Labor Statistics (BLS). Productivity is one of the modifiable parameters, and is highly subjective to estimate for the future. It was re-calibrated in our (contracted) revision of the model's "black box."

Dental Insurance. Data on percent of population insured for dental expenses is from the Health Insurance Association of America (HIAA), and these data are congruent with those used by the NIH's National Institute of Dental Research (NIDR).

Other Data Sources (Government). EMODS' other variables, primarily but not exclusively economic and demographic data, are updated annually from Government sources, some of which include projections or forecasts as well; for other variables, projections are generated by the EMODS model itself. These Government sources include the U.S. Census Bureau, National Center for Health Statistics (NCHS) for updates of number of dental visits, and the National Institute of Dental Research (NIDR) for figures on numbers of decayed and filled teeth.

Physicians

Current Models. Our two current physician models are the physician supply model and the physician requirements model. The American Medical Association (AMA) provides data from their Master File for our aggregate, specialty and State physician supply models. In addition, data from the AOA (American Osteopathic Association) provides comparable data on doctors of osteopathic medicine (DOs). MDs and DOs (doctors of allopathic and osteopathic medicine), when combined, total to all physicians. Some AMA and AOA data are used in the base of the physician requirements forecasting model. Similarly to the dental data base, educational data for MDs and DOs (applicants, enrollments, graduations), are obtained from the Association of American Medical Colleges (AAMC) and American Association of Colleges of Osteopathic Medicine (AACOM), respectively.

Physician Supply Model. The physician supply model projects total supply, specialty supply, and State supply through the year 2020.

The total (aggregate) submodel projects MD and DO physicians by gender, age, and country of graduation. The specialty submodel distributes the projected total supply of Mds by principal specialty (for 37 specialties and subspecialties), and by place of activity (office-based, hospital-based residents and staff, and other). The State submodel distributes the projected total supply of MDs by State and Census region for four major specialty groups, by activity, gender and county of graduation.

Disaggregation of DO data is more limited because of data limitations, although DO projections are made at the State level.

Physician Requirements Model. The physician requirements model is a demographic utilization model which provides forecasts of utilization-based requirements for primary and nonprimary care physicians.

Variables Used. Data inputs for the supply model include MDs by gender, age, (self-designated) specialty, location (State of practice), place of activity (office- or hospital-based practice, or "other"), and country of graduation; DOs by age, gender, and specialty; first-year student enrollment; and death and retirement rates. Due to the detailed level of these data, the Bureau has a specialty and subnational model, as well as an aggregate model. These data are purchased under a contract from the AMA's physician masterfile. DO forecasts are made within the MD specialty forecasting model. However, we have considerably less specialty detail for DOs, redistricting the range of our specialty model when we include DOs. The Bureau is currently working with the AOA to improve our forecasts for DOs, in conjunction with the AOA's own efforts to improve its own masterfile (see our discussions with the AOA in early 1991, as mentioned earlier).

The demographic-utilization-based requirements model contains a somewhat different mix of physician activity (service) categories, each of which has associated utilization rates for ten age/sex categories. Most of the current utilization figures come from survey data provided by CDC's National Center of Health Statistics (NCHS), but others, such as hospital admission and outpatient visit trends, are from the AHA and other non-governmental agencies. The AMA provides their own analyses of physician supply[3] and utilization, which we and others regularly compare to our results.[4]

Physician Geographic Distribution Models. The Bureau has also two geographic distribution models, revised in the late-1980s — for total and primary care physicians, to assess the geographic diffusion of physicians. These are econometric models of physician supply (MDs and DOs, by county, by age), as it relates significantly to population, income, and employment (total model) and to hospital admissions (primary care model).[5]

Hospital Personnel

One of the main data sources for many categories of health personnel is the annual survey by the AHA, since many of them are based in the hospital setting.. This is especially the case for the allied and associated health professions, where only limited association data are available for analysis.

AHA Surveys. We utilize American Hospital Association data from three AHA surveys:

Our main AHA database is Schedule G ("Personnel") from their annual survey. Based on the data tape we prepare an annual report, the latest of which is ***Hospital Personnel Trends 1981-1988***.[6] A new report (through 1989) is now in process. These report contain an analysis and data on 36 personnel categories, for

all registered hospitals and community hospitals, and for total personnel and FTE personnel, by Census Region, State, and bedsize. Our referenced latest report of these AHA data investigates overall hospital employment trends, trends by occupation, region, and type of hospital.

Another AHA data base we use are the Hospital Panel Survey data, which we receive on a monthly basis. These contain data on utilization, on hospital revenues and expenses and some minimal data on manpower. These reports compare the same month from year to year (e.g., December 1990 to December 1989). In our referenced report we use information from this survey to introduce general trends in the hospital sector.

A third AHA survey, relatively new but of great use to our analytical activities, is the Survey of Human Resources. This annual survey asks questions concerning recruitment, retention, and time required to fill a vacancy — which positions are hardest to fill, which take longest.

Trends Report. Our 1990 publication, ***Trends in Hospital Personnel 1981-1988***, emphasizes the occupational profile of hospitals, and presents statistical tables and analysis for 36 occupations, by Census Region, State and bedsize. An updated report is currently in preparation.

Area Resource File (ARF)

We enter many of these data sets into the Area Resource File (ARF), which is our (publicly available) in-house depository for selected variables from our various databases. We sometimes use the ARF files directly for various models, papers, and reports; at other times we excerpt selected variables from ARF and build our special databases as needed, often combining with other special files.

The ARF file is available to, and in much demand, by many "outside" users, such as researchers and analysts at universities, private organizations, State and local governments.[7]

Conclusion. The analysts in the Bureau of Health Professions are dependent on secondary data sources to a large degree. We obtain these from both public sources and agencies, such as the Census, Bureau of Labor Statistics (BLS), National Center for Health Statistics (NCHS), or Health Care Financing Administration (HCFA) and from associations. The cooperation of these associations and their willingness not only to sell us their data sets, but to listen to and consider our input and data requirements is essential to the success of our efforts. There is and active and ongoing exchange between us and "them," which benefits both sides.

References

Dentistry

Bronstein, G.L., Clark, N.L., and Traxler, H. "Econometric Model of the Dental Sector: Epidemiological, Economic, and Computer Factors Involved in Revising and Updating a Forecasting Model." In ***Federal Forecasters Conference 1990: Proceedings*** (# AGES 9109), Washington, D.C., March 1991.

Congress of the United States, Congressional Budget Office. ***The Economic and Budget Outlook: Fiscal Years 1992-1996.*** Washington, D.C., January 1991.

Eklund, S.A. and Roehrig, C.S. "Accounting for structural changes in dental epidemiology: revision of the Econometric Model of the Dental Sector." Prepared for presentation at annual conference of the American Public Health Association, New York City, Oct.-Nov. 1990. Draft, September 1990.

Feldstein, P.J. and Roehrig, C.S. "A National Econometric Forecasting Model of the Dental Sector." ***Health Services Research*** 15(4):415-432, Winter 1980.

Furino, A., Douglas, C.W. "Balancing Dental Service Requirements and Supplies: The Economic Evidence." ***Journal of the American Dental Association*** 121:685-692, December 1990.

Hixson, J.S. "Long-Term Forecasts of Economic Activity in the Dental Sector." ***Journal of the American College of Dentistry*** 48(2):78-85, Summer 1981.

Traxler, Herbert. "The Econometric Model of the Dental Sector (EMODS)." Outline of presentation at the Federal Forecasters Conference, September 6-7, 1989, Washington, D.C.

U.S. Department of Health, Education, and Welfare. ***The Economic Model of the Dental Sector: Purpose, Scope, and Uses.*** OMA Report No. 80-44. U.S. Government Printing Office, Washington DC, March 1980.

U.S. Department of Health and Human Services, Health Resources and Services Administration. ***Econometric Model of the Dental Sector (EMODS): Summary of Work Performed.*** VRI-HRSA-9 FR90-2. July 26, 1990.

U.S. Department of Health and Human Services, Social Security Administration. ***1990 Annual Report of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.*** Washington, DC, April 19, 1990.

U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions. ***Seventh Report to the President and Congress on the Status of Health Personnel in the United States.*** DHHS No. HRS-P-OD 88-1. U.S. Government Printing Office, Washington, DC, March 1990.

U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions. ***Eighth Report to the President and Congress on the Status of Health Personnel in the United States.*** Draft, August 1991.

Hospital Personnel

US DHHS, HRSA. ***Trends in Hospital Personnel, 1981-1988.*** Rockville, Md., DHHS Publication No. HRS-P-OD-90-4, 1990.

US DHHS, HRSA. Updated Trends in Hospital Personnel (title pending). In preparation, 1991.

Physicians

Cultice, James. [Sept. 6, 1990 presentation at FFC ...]

US DHHS, HRSA. [Final Report by contractor ... 1985 ...]

7th Report to Congress: see Dentistry above.

8th Report to Congress: see Dentistry above.

Footnotes

1 Bronstein, G.L., N.L. Clark and H.G. Traxler, "Econometric Model of the Dental Sector: Epidemiological, Economic, and Computer Factors Involved in Revising and Updating a Forecasting Model," ***Federal Forecasters Conference 1990: Proceedings***, (U.S. Department of Agriculture, Economic Research Service, AGES No. 9109), Washington, D.C., March 1991. An overview of EMODS is given in Traxler, H., "The Econometric Model of the Dental Sector (EMODS)," Outline of presentation at the 1989 Federal Forecasters

Conference, Washington, D.C., September 1989 (available from the author).

[2] Furino, A. and C.W. Douglas, "Balancing Dental Service Requirements and Supplies: The Economic Evidence," ***Journal of the American Dental Association*** 121:685-92, December 1990.

[3] The latest ***Projections of Physician Supply: 1990 Update*** by the AMA Center for Health Policy Research was revised in December 1990. It reports on such issues as improved data on foreign medical graduates (FMGs), "which would explain why both the AMA's projections and those of the Bureau of Health Professions have tended to underestimate the size of the FMG population." (page 2).

[4] Singer, A.M. "Projections of Physician Supply and Utilization by Specialty'" ***Academic Medicine***, May 1989, p.219

[5] Traxler, H.G. "The Influence of Selected Economic Variables on the Future Geographic Distribution of Physicians in the U.S.," paper presented at the annual meeting of the American Economic Association, Atlanta, GA, December 1989 (under revision, available from the author).

[6] U.S.DHHS, Public Health Service, HRSA, Bureau of Health Professions, ***Trends in Hospital Personnel: 1981-1988***, DHHS Publication No. HRS-P- OD-90-4, November 1990.

[7] For a description of the area resource file see U.S.DHHS, Public Health Service, HRSA, Bureau of Health Professions, ***The Area Resource File (ARF) System: Information for Health Resource Planning and Research***, ODAM Report No. 7-89 (available from the author).

Annual Precipitation Estimates — 1 to 4 Years in Advance

Charles A. Perry, U.S. Geological Survey, WRD, Lawrence, Kansas 66049

Precipitation forecasts have been limited to several days by numerical models of the atmosphere and to a few months by statistical methods [1,2]. Forecasts beyond these time frames have been only speculative and are based on extrapolation of established trends or periodicities. A significant relation between solar irradiance and regional precipitation is shown here. Variations in irradiance can affect the amount of energy that is injected into the world's tropical oceans. Increases or decreases of absorbed solar-energy are later manifested as increased or decreased regional precipitation. A lag time, unique to each region in the United States, occurs between irradiance variations and regional precipitation. The lag times are dependent upon the distances and the major ocean-current velocities from areas in the tropical oceans that absorb the majority of the solar energy to areas of the ocean that supply moisture directly to the regions in question.

Earth-satellite measurements since 1978 have indicated that total solar irradiance has a long-term variation of at least 0.1% during the 11-year period of a sunspot cycle[3]. However, variations greater than 0.1% have been noted in consecutive monthly irradiance averages. The solar-irradiance record has been computed prior to 1978 by an empirical model developed by Foukal and Lean[4]. The variations in the modeled solar irradiance during the last four sunspot cycles (1945-88) show a significant correlation with regional precipitation. The variations in irradiance are small, on the order of 1 W/m^2 per month, so their effect needs to be amplified to cause significant climatic variations.

One possible medium for amplification of the solar-irradiance variations may be through the transfer of energy from the oceans to the atmosphere. Variations in the temperature of the ocean surface, specifically sea-surface temperatures (SST), have been linked to atmospheric-pressure anomalies[5]. There is evidence that an anomalously cool SST in the eastern Pacific was responsible for the severe 1988 North American drought[6]. Ocean currents serve as the major conveyors of energy from the tropics toward the poles. The mechanism proposed for the coupling of solar irradiance with regional climate through the ocean consists of three components. These are: (1) absorption of solar energy by the transparent tropical oceans in a deep surface layer, (2) transport of that energy from the tropics to the temperate regions by major ocean currents, and (3) transfer of that energy into the atmosphere by evaporation processes which supplies moisture and energy to low pressure systems that cause precipitation.

A difference of only 1 W/m^2 reaching the Earth's surface and penetrating the ocean can be translated into a measurable change in ocean temperature. Approximately 75% of the power reaching the Earth's surface is absorbed in the top 10 m of clear ocean water[7]. However, blue light can penetrate to nearly 100 m. If a 1 W/m^2 irradiance difference persists over a time span of 1 year and if it is assumed that 75% of that difference in energy is absorbed by the top 10 m of the ocean (the remainder absorbed below 10 m), then that 10 m column of ocean water could have a temperature variation of more than 0.5 °C. Lewis et al[8] show that solar radiation penetrates to a significant depth (instead of being absorbed at the surface) in the clear waters of the western tropical Pacific Ocean. This explains the discrepancy between observed SST in the western Pacific and those predicted by current ocean-atmosphere models. By occurring over an area of several thousand square kilometers, this variation in the total energy of a very large volume of water could have an affect on the atmosphere above it for a considerable period of time.

The Pacific Gyre and its minor circulations are the conveyors of absorbed solar energy from the central and western tropical Pacific to locations north and east. If incoming solar energy varied on time scales of months to years, different parts of the gyre would receive varying amounts of energy as water moved through the tropics in its journey around the gyre. During a period of decreased solar irradiance, a part or pool of the tropical ocean would receive less energy and become anomalously cool, whereas increased irradiance would result in an anomalously warm pool. These pools of warmer or cooler water are drawn around the Pacific Gyre like riders on a carousel. A lag time of several years could elapse between the time that energy is absorbed into the tropical ocean and the time that the energy is finally released to the atmosphere. By then the pool of ocean water may have traveled thousands of miles from its initial location. Other factors such as cloud cover, atmospheric turbidity, latitude, wind speed and ambient air temperature affect SST, but irradiance variations may play an important role in ocean temperatures below the surface.

Evaporation from the surface of the ocean is the mechanism that may have the greatest amplification of the effect of solar-irradiance variations on climate. The vapor pressure of water increases by about 6 to 7% for each 1°C of increase in temperature between 5 and 25 °C [9]. For example, a +2°C anomaly in SST has nearly 28% more water vapor above it than does a -2°C anomaly. This increase in water vapor could significantly affect precipitation by increasing atmospheric moisture fields from which further amplification of irradiance variations could occur through the dynamic atmospheric processes of storm and precipitation formation.

Annual totals (January to December) of monthly solar irradiance differences were computed from the monthly modeled irradiance values. Correlation coefficients between these annual totals and annual State-divisional precipitation values for the 344 regions in the United States for the years 1950-88 were computed with lag times of 0 to 7 years. The different lag times were used to examine the time of transport of stored solar energy in the ocean water from the tropical Pacific to the atmosphere over locations in North America. Correlation coefficients (R) that ranged between +0.65 and -0.51 occurred in patterns that can be explained by upper atmospheric wind fields and could result from ocean-temperature anomalies. To be significant at the 1% level, R must be greater than 0.37 or less than -0.37. These patterns of similar correlation coefficients showed a tendency to migrate across the United States from west to east at a rate nearly equal to the velocity of the North Pacific Drift Current (approximately 8 km/day). The highest positive correlations occurred in the Pacific Northwest States (Washington, Oregon Idaho, western Wyoming, and the northern regions of Utah, Nevada, and California) with a lag time of 4 years. This lag is approximately equal to the time for water to travel from the western tropical Pacific Ocean to the Gulf of Alaska within the Pacific Gyre. Of the 19 regions in Washington and Oregon, 13 had correlation coefficients greater than 0.50, demonstrating the consistency of responses of neighboring regions. With a 4-year lag time between irradiance changes and precipitation, droughts (dry high-pressure development) in the Pacific Northwest coincided with periods of negative irradiance differences, and wet periods (moist low-pressure development) coincided with periods of positive

differences (Fig.1). Other groups of regions in the United States demonstrated other significant lag times consistent with their locations.

The persistent drought in the Pacific States from 1984-91 can be traced to a period of decreasing solar irradiance that occurred between 1980-87 (Fig. 1). Projections utilizing the correlation between solar-irradiance variations and Pacific Northwest precipitation indicate a period of greater than average precipitation in the Pacific States beginning 4 years after the substantial increases in solar irradiance that occurred in 1987-89.

Footnotes

1. Suranjana, S. and Van Den Dool, H.M. M. *Wea. Rev.* 116, 2522-2526 (1988).

2. Livezey, R.E. *Bul. Am Meteor. Soc.* 71, 300-309 (1990).

3. Willson, R.C., and Hudson, H.S. *Nature*, **332**, 810-812 (1988).

4. Foukal, P. and Lean, J. *Science*, **247**, 556-559 (1990).

5. Wallace, J.M., Smith, C. and Jiang, *Q, J. Climate*, **3**, 990-998 (1990).

6. Palmer, T.N. and Brankovic, C. *Nature*, **338**, 54-57 (1989).

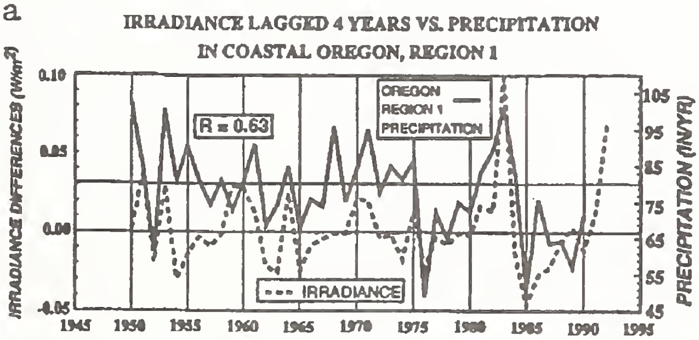
7. Beer, T. *Environmental Oceanography*, **206** (1983).

8. Lewis, M.R., Carr, M., Feldman, G., Esaias, W. and McClain, C. *Nature*, **347**, 543-545 (1990)

9. Weast, R.C. *CRC Handbook of Chemistry and Physics*. D-168 (1982).

ACKNOWLEDGEMENTS. I thank J. Lean for providing the modeled solar irradiance data and G. Reid for critical comments on a preliminary manuscript.

a



b.

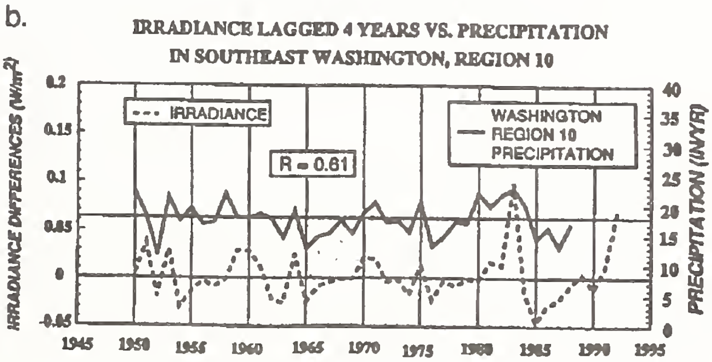


Figure 1. Comparison of annual total of monthly solar-irradiance differences lagged 4 years and annual precipitation for (a) coastal Oregon, Region 1 and (b) southeast Washington, Region 10 (R=correlation Coefficient)

Forecasting Grain Losses in Chinese Agriculture from Natural Hazards — A Kalman Filter Approach

HaoQuan Chu, Arthur D. Little, Inc.

I. Introduction In empirical economic forecasting, researchers often have to deal with the problem of changing model structures during the time of observation. Aggregate or historical data may also include measurement errors. When both problems are present, standard regression analysis does not generate satisfactory results since some of the underlying assumptions of the standard model are violated. Specifically, the assumption that the coefficients of the regression model are constant over time may not be appropriate.

One example cited in this paper is an analysis of the effect of natural hazards on agricultural yield in China. The data on natural hazards in China are not well defined, and the technological and political environment changes over the years as well. Under these conditions, we should consider the changing coefficient regression model (Chow, 1984) which treats structural coefficients as time-varying random variables. The advantage of the model is that the problem of noisy data and changing structure can be handled within the parametric regression analysis framework.

The Kalman filter is used in this paper as the main technique in our changing coefficient regression model. It is shown that the computational burden is much less when the E-M algorithm is applied to Kalman filter.

The plan of the paper is as follows. Section two presents the econometric model; section three provides some background on Chinese agriculture and natural hazards; section four describes the data set; section five presents the regression and forecast results; the final section gives some concluding remarks.

II. The Econometric Model Among the changing coefficient regression models, the Kalman filter (Kalman, 1960) has been widely used in engineering and financial analysis areas for smoothing and forecasting purposes since it has the advantage of allowing inaccuracy in variables *and* structural change over time. In a linear regression framework, the Kalman filter can be represented in the state-space notation:

(1) $y_t = x_t \beta_t + v_t,$

(2) $\beta_t = M\beta_{t-1} + w_t, \qquad t=1,\dots,T.$

The first equation is called the measurement equation, where y_t is the dependent variable, x_t is an $1 \cdot K$ vector of explanatory variables, v_t is the i.i.d. disturbance term which is assumed to have zero expectation and a covariance matrix R. The second equation, which defines the path of the stochastically evolving structure, is called the transition equation, where M is a $K \cdot K$ matrix of fixed coefficients and w_t is a $K \cdot 1$ vector of independent normal disturbances which have zero expectations, a covariance matrix Q, and is uncorrelated in different time periods.[1] Finally, the initial value of ß is assumed to be normally distributed with mean • and variance matrix Σ.

Suppose the parameters of the evolving path (R, M, Q) are known, then the problem is to estimate β_t conditional on the accumulated observations from time period 1 to t. In some textbooks, the estimation procedure may look sophisticated. However, the model can be interpreted as generalized least squares (GLS) estimation (see Harvey and Phillips, 1982). Cuthbertson (1988) gave a detailed description on how to apply the Kalman filter using standard GLS techniques.

The major difficulty with the GLS method is that the variances and the covariance matrixes of the model are usually unknown. There are a few methods suggested in the literature on how to estimate the variances (see, for example, Cuthbertson, Harvey and Phillips). Compared with the methods of ad hoc assumptions and the sensitivity analysis using "plausible" values, the maximum likelihood estimation (MLE) is theoretically more appealing since it simultaneously estimates the variance-covariance coefficients and system parameters. However, common ML methods like Newton-Raphson techniques generally involve the calculation of the inverse of the second-order partial matrix which can be very tedious computationally. An exception is the EM algorithm (Dempster et al 1977) applied to the ML estimation of the Kalman filter, which was pioneered by Shumway and Stoffer (1982). The EM approach does not require the calculations of second partial derivatives; it also has the advantage that the value of the likelihood is guaranteed to converge to a stationary point. This stationary point is at least a local maximum, if not a global maximum point, so once the algorithm is implemented, finding the global maximum point through the search of a reasonable range of initial value is relatively easy.

The joint log likelihood of the complete data set, as shown by Shumway and Stoffer (1982), can be written as

$$\log \mathcal{L} = -\frac{1}{2}(\log|\Sigma| + (\beta_0-\mu)'\Sigma^{-1}(\beta_0-\mu) + T\log|Q| + \sum_{t=1}^{T}(\beta_t - M\beta_{t-1})'Q^{-1}(\beta_t - M\beta_{t-1})$$
$$+ T\log|R| + \sum_{t=1}^{T}(y_t - x_t\beta_t)'R^{-1}(y_t - x_t\beta_t)),$$

where log L is to be maximized with respect to •, Σ, M, R, and Q. The EM algorithm is an iterative procedure which can be implemented through two steps: the first step is *expectation*, where the expected value of the log likelihood with respect to the observed dependent variable is calculated:

$$D(\mu, \Sigma, M, Q, R) = E(\log \mathcal{L} \mid y_1, \dots, y_T),$$

the second step is *maximization*, which maximizes function D with respect to the parameters. Based on the conditional expectation of the log likelihood, Shumway and Stoffer introduce a set of formulas which can be used recursively in calculating the expected values of ß and its covariances which are defined as:

$$\beta_t^{s} = E(\beta_t \mid y_1, \dots, y_s),$$
$$P_t^{s} = \mathrm{cov}(\beta_t \mid y_1, \dots, y_T),$$
$$P_{t,t-1}^{s} = \mathrm{cov}(\beta_t, \beta_{t-1} \mid y_1, \dots, y_T).$$

where s=1,...,T. They also provide the maximizing formula for ß, M, •, Σ, R, and Q based on calculated expectation values. So for empirical economists, Kalman filter estimation using the E-M algorithm involves just three steps:

(1) Expectation: calculate ß, M, •, Σ, R, and Q using the recursive equations.

(2) Maximization: update the estimates of ß, M, •, Σ, R, and Q using the maximizing formula and calculate the value of the log likelihood.

(3) Repeat step (1) and (2) until the maximized values converge.

In section five, this simple algorithm is applied to data on Chinese agriculture and natural hazards.

III. Chinese Agriculture and Natural Hazards Typical of a developing country, agriculture has been one of the most important sectors in the Chinese economy. In the past 40 years, Chinese

agriculture has experienced a lot of policy and technological changes; yet despite these changes, natural hazards such as floods, typhoons, and drought remain key factors in the agriculture yield.

Due to geographical diversity, each year there are always some areas affected by the natural hazards that cause harvest losses or even no harvest at all (see Table 1). In the early and middle summer of 1991, the *huangmei* season caused an extraordinarily large amount of precipitation that flooded much of the Yangtze River area. Over 1,000 people died and millions were without homes or food. Due to severe shortages, the vegetable prices in the big cities like Shanghai increased more than 200%; summer crop grain loss is estimated at 4.35 million tons for Anhui Province alone.

Because of its huge population (around 1.2 billion now) and relatively limited sown area, each year China needs to import grain from other countries such as the United States, Canada, and Australia. It is thus important to be able to forecast the grain loss caused by natural hazards. Unfortunately, there are few quantitative studies in this area. The most recent one that can be found is by Kueh (1984). Kueh analyzed the annual data for average grain yield and natural hazards from 1952 to 1981 and performed simple correlation regressions. The Cultural Revolution disrupted the data collection process, so the 1967-69 data are unavailable, which makes it natural to assume that there are two structurally different subperiods (1952-1966 and 1970-1981). It was found that in the post-Culture Revolution years, the average grain yield became more weather-resistant. This paper updates Kueh's time series and attempts to improve his estimation results with more flexible Kalman filter techniques.

Description of Data Set The principal data needed are those for agricultural output and natural conditions. The official Agriculture Statistical Yearbook classifies farm crops into three major categories: grain crops, industrial crops and other farm crops. Like Kueh, only the grain crops are studied in this paper since (1) grain crops consist of approximately 80 % of the total sown area; and (2) grain is the major agricultural product that China imports. It is also more reasonable to link total grain production to the aggregate measure of natural hazards since five grain crops[2] cover almost all geographical areas and are grown around the year.

Kueh primarily used the official natural hazard data in his analysis. Calling the data a weather index, he made comparisons with the categorical weather index created by Tang (1980). However, it should be noted that the term "natural hazard" means more than just weather: plant diseases, pests, and rats also cause great damage and they may not be related to the weather. The highly aggregated official natural hazard data included *shouzai* (affected) area and *chengzai* (heavily affected) percentages among the *shouzai* area. The *chengzai* area is defined as an area that suffers 30 percent or more loss in the harvest from natural disasters.[3] Although a lot of published data in China serve propaganda purposes, current official hazard data should be reliable (see Kueh for more details).

Shouzai and *chengzai* information can be combined into one variable. However, the aggregation is avoided here since strong assumptions have to be made regarding the exact percentages of crop loss in the heavily and "lightly" affected areas, which may vary from year to year. Even if this analysis does not make such assumptions, it should be kept in mind that *shouzai* and *chengzai* information only partially reveals the true magnitudes of natural disasters.

During the observed time period (1952-88), there are many drastic policy changes which could have strongly influenced the production structure. Unfortunately, these factors are very hard to quantify. In the meantime, production technology changes are also very significant. These include increases in mechanization and irrigation, multiple-cropping, and greater applications of chemical fertilizer and pest control. In the literature of production function estimation, overall "technical change" is often represented by a time trend. This study uses a single variable, total machinery horsepower, to represent the technical change. Admittedly this variable is far from perfect, but it is perceptibly better than the time trend. Other data, such as chemical fertilizer use, are not included because of the problem of multicolliearity.

Finally, the data for total and grain sown areas are gathered. After 1970, total sown area declines slowly over the years due to increasing demand for industrial use and residential house construction. The percentage of total sown area used for grain crops also gradually declines from around 85% in the 1950's to around 76% in late 1980's because of increasing demand for industrial crops[4] and other farm crops. Nevertheless, grain output increases significantly, a fact that should be mainly attributed to the technical change.

Tables 1 - 3 show the general time trend of the major variables. The other input data, such as the quantity of capital and labor, are hard to find. Therefore, it is assumed that labor input per hectare is constant over time and capital input is represented by the total machinery horsepower.

Estimation Result Initial regression estimations use unit grain yield as the dependent variable and percentages of affected and heavily affected sown area as two explanatory variables, with a time dummy representing technical change. Technical change turns out to be a dominant factor; the effects of natural hazards were highly insignificant, even though the signs are negative. It is clear that the fluctuation in unit yield is not significant enough to be explained by the variations in the degrees of natural hazards. On the other hand, Kueh's regression does not incorporate the effect of technical change and is based on the variables expressed in their deviation from the trend. He reports a correlation coefficient between grain yield and the weighted *shouzai* area index of only 65% over 1952-81 and 58% over 1970-81. As a third alternative, unprocessed raw data can be used in specifying the functional relation. It is assumed that the annual total grain yield (G) is a linear function of total sown area (A), *shouzai* area (S), *chengzai* percentage (C), and total machinery horsepower (H)[5]:

$$G_t = a_0 + a_1 A_t + a_2 S_t + a_3 C_t + a_4 H_t + e_t, \qquad t=1,...,T. \tag{3}$$

Table 2 shows OLS regression results based on this simple model. The 1952-88 data give a very good fit (R^2 = .96). The coefficients for all variables except the intercept are statistically significant. The insignificance of the intercept implies that there will be no grain without land. Since the data cover 37 years and have a 3-year break in the middle which represents a clear policy change, it is natural to run regressions for the two subperiods (see Table 2).[6]

The regression coefficients imply expected effects of the four variables on the grain yield: over all the time periods (1952-88), machinery horsepower and land have strong positive effects on grain output, while the areas affected by natural hazards and the percentage of heavily affected areas have a negative impact on grain yield. By comparing the regression coefficients of the two subperiods, it can be seen that over time the grain yield depends more on machinery horsepower and less on land, as the coefficient

of sown area for the second subperiod becomes insignificant. The grain's resistance to natural hazards also increased, as the coefficient for *chengzai* percentage becomes insignificant in the second subperiod. Production changes such as multicropping, and scientific research on major crops certainly helped. The data fit considerably better in the second subperiod, probably reflecting the combined effects of improvements in the data collection system and the stabilization of grain production.

Table 2 also presents separate regression results for the two most important grain crops: rice and wheat. Note that rice is grown mainly in the south and wheat is grown in the north. So using rice or wheat yield as dependent variables is not strictly consistent with the independent variables which cover the whole country. It is a little surprising that results are quite similar with the grain equation. Nevertheless, there are two major differences. First, the coefficient for the *changzai* area for the wheat equation is insignificant for all three regressions. It is interesting to see whether wheat areas are rarely heavily affected by natural hazard. Second, for both rice and wheat equations all coefficients except machinery horsepower are insignificant in the 1970-1988 subperiod. Attempts are made to disaggregate the grain sown area into individual grain crop sown areas. However, the regression results (not shown here) are less than satisfactory.

The OLS regression results in Table 2 provide us with valuable information on the relationship between the grain yield and the influential factors. However, because of obvious structural changes and data noise, this set of results is not expected to serve well for forecasting purposes. In other words, the assumption of fixed coefficients prevents us from catching the unobservable structural changes over time. Also, data noise and structural instability in the earlier time periods are likely to have an undesirably large impact on the regression coefficients. So the changing coefficient model is estimated using Kalman filter. The model can be expressed in matrix form:

(4) $G_t = X_t B_t + v_t,$

(5) $B_t = MB_{t-1} + w_t, \qquad t=1,\ldots,T,$

(6) $B_0 = (X'X)^{-1}X'G,$

where $X_t = (1\ A_t\ S_t\ C_t\ H_t)$, $X = (X_1'X_2' \ldots X_n')'$ and others are defined the same as equations (1) - (3). Equation (6) implies that initial values for B are estimated by OLS regression using the initial n time periods of data. The initial covariance matrix for B is also obtained from the OLS regression. This method has been suggested in the literature along with some other methods to obtain the initial value. Shumway and Stoffer use ad hoc guessing of the initial value. However, with the current framework of four independent variables it is fairly difficult to make a good guess.

Table 3 presents estimation results of the Kalman filter. Two subperiods (1952-66 and 1970-88) are combined as if they are a continuous time series. The difference between the two sets of Kalman filter results is the time period used to obtain the initial B_0 estimates. B_0 of MLE(1) is estimated by setting n=T, while in MLE(2)'s case n=5. The results show little difference although they are quite different from the OLS results, taken from Table 2. When the coefficients are allowed to vary over time, the technological structure, represented by total machinery horsepower, turns out to be more stable over time. The effect of land over time seems to increase; and the effect of natural hazards tends to decrease over time. These general trends are consistent with the finding by Kueh and also consistent with OLS results in Table 2.

The 1990 Statistical Yearbook was published recently and contains 1989 agriculture data. For 1989, total sown area is 146.55 million hectares, total machinery horsepower is 361.3 million hp. and *shouzai* area is 47.0 million hectares with 52.0 *chengzai* percentage. These values are substituted into the estimated structural equation to predict the 1989 grain yield, which is 40.76 million tons. Table 3 shows the forecasted values and forecast biases for three regressions. It is clear that the MLE regressions far outperform the OLS regression.

One obvious reason that OLS performs so poorly is that the regression coefficients are heavily influenced by the early time period data, which are noisy and structurally different than that of 1989. One might think OLS results from the second subperiod would give better forecast results. Table 4 confirms this hypothesis; the forecasting bias for the second subperiod OLS is reduced from 5.56 to 3.29 (10 million tons). Table 4 also shows the MLE results and forecasts using the same data set (1970-88) for both initial B_0 estimation and the Kalman filter. MLE produces a forecasting bias of 0.82, still far better than OLS.

Comparing Tables 3 and 4, it is interesting to see that if one wants to stick with simple OLS techniques, then one should exclude the data which are very noisy or structurally different from these of the forecasting period. However, the same logic does not apply to the changing coefficient model, since it not only absorbs the noise and changing structure, its forecasting accuracy actually gained a little by using all available information.

Rice and wheat yield forecasts using the changing coefficient model and OLS regression produce similar result. The study can also be extended to forecast the yield of other agricultural sectors such as industrial crops, which are becoming more and more important.

Finally, it is important to keep in mind that a changing coefficient model may not always outperform a fixed coefficient model by such a large margin. Certainly not all data sets will have the phenomenon of declining structural instability and declining measurement error over time.

Concluding Remarks In an attempt to forecast annual grain yield in China, this paper shows that the Kalman filter can be a powerful tool in empirically coping with problems of unobservable effects that alter production structure over time. Although OLS regressions are satisfactory in finding the major trends and the direction of the effects of influential factors on grain yield, it is demonstrated that we need to relax the restriction on regression coefficient to "update" the structural coefficient in order to produce accurate forecasts. Although the technique for estimating the changing coefficient model is more complicated than OLS technique, judging from the forecasting gain it is well worth the effort to try the MLE.

The Chinese agriculture time series in this paper actually has three missing observations. However, this paper treats the time series as if there are no missing data. It is interesting to see how the regression results and prediction power will improve if the methodology suggested by Harvey (1989) is adopted.

One of the potential applications of this analysis is policy evaluation. The natural hazards may not be all "natural". Actually the *shouzai* area can be disaggregated into pure natural disasters and largely man-made hazards (Kreimer and Zador, 1989). Knowing the magnitude of the man-made hazards is very important to quantitatively measure the actual loss due to policy mistakes in Chinese history. It is unfortunate that this type of data is not yet available.

In recent years more detailed data on Chinese agriculture

has become available. In both the ***China Statistical Yearbook*** and ***China Agriculture Statistical Yearbook***, the *shouzai* and *chengzai* information is now available at the provincial level. A panel data estimation should give much more insight on the model structure, which is changing quickly with technology and policy.

REFERENCES

China Agriculture Statistical Yearbook, 1981-1989, Agriculture Publishing House, Beijing, China.

China Statistical Yearbook, 1980-1989, Chinese Statistical Press, Beijing, China.

Chow, G.C. (1984), "Random and Changing Coefficient Models," in ***Handbook of Econometrics***, Vol. 2, chapter 21, pp. 1213-45.

Colloquium on Disasters, Sustainability and Development: A Look to the 1990s, (1989) Kreimer, A. and M. Zador, eds. Environment Working Paper No. 23, World Bank.

Cuthbertson, K. (1988) "Expectations, Learning and the Kalman Filter," ***The Manchester School of Economics and Social Studies***, Vol. LVI, No. 3, pp. 223-46.

Dempster, A., N. Laird, and D. Rubin, (1977) "Maximum Likelihood from Incomplete Data via the E-M Algorithm," ***Journal of Royal Statistical Society***, B39, pp. 1-22.

Harvey, A.C., (1989) ***Forecasting, Structural Time Series Models and the Kalman Filter***, Cambridge University Press, Cambridge.

Harvey, A.C., and G. D. A. Phillips, (1982) "The Estimation of Regression Models with Time-Varying Parameters," in ***Games, Economic Dynamics, and Time Series Analysis***, M. Deistler et al. eds., Physica-Verlag, Wien-Würzburg, pp. 307-21.

Kalman, R.E., (1960) "A New Approach to Linear Filtering and Prediction Problems," ***Journal of Basic Engineering***, pp. 35-45.

Kueh, Y. Y., (1984) "A Weather Index for Analyzing Grain Yield Instability in China, 1952-81," ***The China Quarterly***, March, No. 87, pp. 68-83.

Shumway, R. H. and D. S. Stoffer, (1982) "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," ***Journal of Time Series Analysis***, Vol. 3, No.4, pp. 253-64.

Tang, A., (1980) "Trend, Policy Cycle, and Weather Disturbance in Chinese Agriculture, 1952-1978," ***American Journal of Agricultural Economics***, May, pp. 344.

Footnotes

[1] Note that if M is an identity matrix, then the model becomes a random coefficient regression (RCR) model; if we further restrict Q to be 0, then the model reduces to the standard OLS regression model.

[2] These are rice, wheat, corn, soybeans, and tubers.

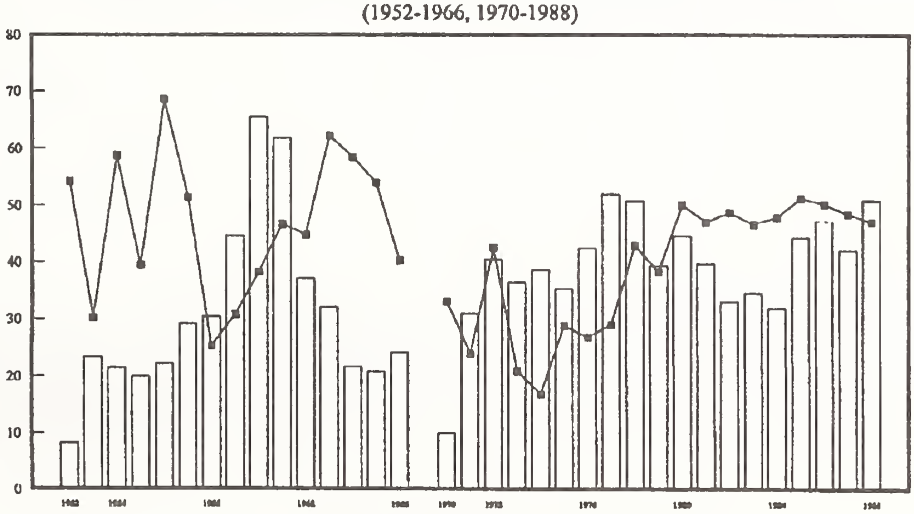
[3] In recent years more detailed data are available for the flood or drought area and for actual amount of crop loss due to different types of hazards.

[4] It includes cotton, peanuts, sugarcane, etc.

[5] Since the number of observations is very limited, severe multicollinearity problems prevent us from adopting more flexible functional forms.

[6] The situation is so obvious that Chow's structural test is unnecessary.

GRAPH 1: SOWN AREA AFFECTED BY NATURAL HAZARDS
(1952-1966, 1970-1988)

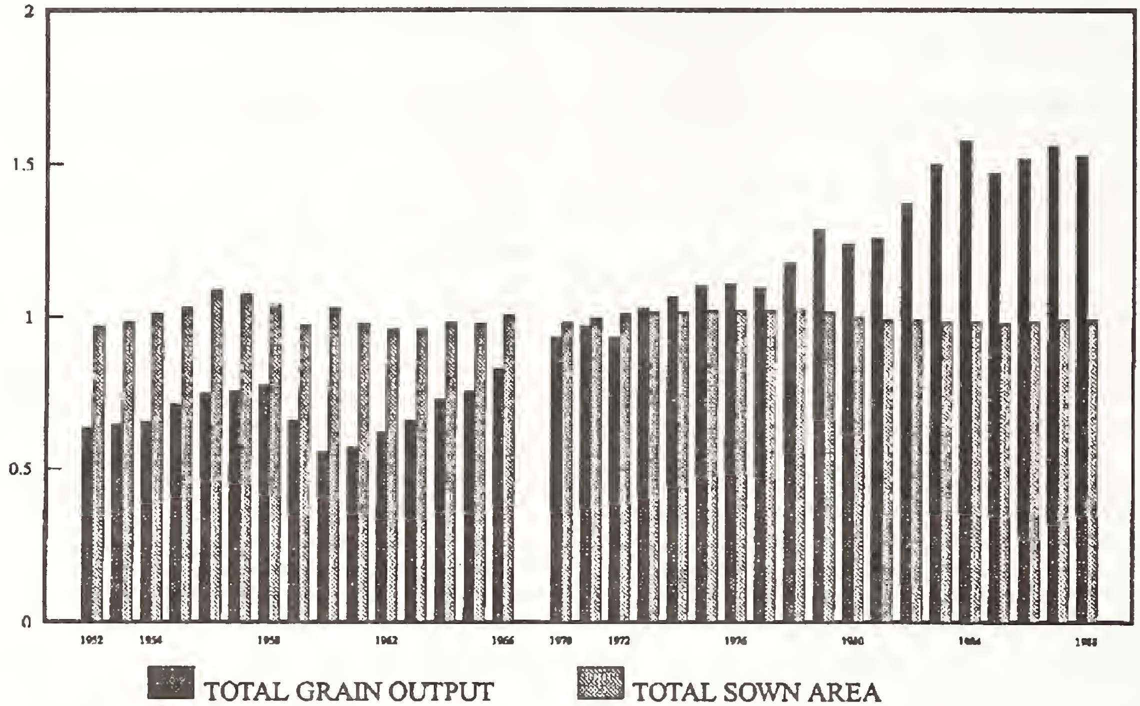


NATURAL HAZARD AREA (MILLION HECTARES)

PERCENTAGE OF HEAVILY AFFECTED AREA

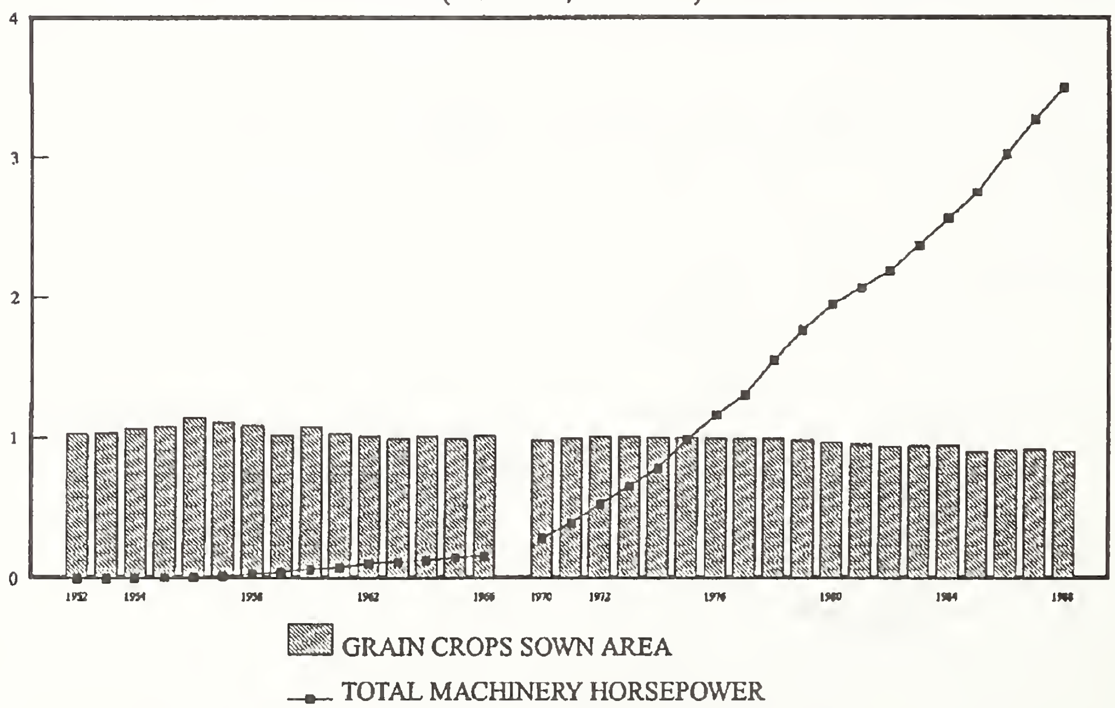
Note: Heavily affected implies a grain loss of more than 30%.

GRAPH 2: TOTAL SOWN AREA AND GRAIN OUTPUT
(1952-1966, 1970-1988)



Note: the values are normalized by the sample means.

GRAPH 3: TOTAL MACHINERY HORSEPOWER AND GRAIN CROPS SOWN AREA
(1952-1966, 1970-1988)



Note: The values are normalized by the sample means.

TABLE 1. CHINA: AREAS AFFECTED BY NATURAL HAZARDS FOR SELECTED YEARS

YEAR	1978	1980	1984	1985	1988
(1) AREAS AFFECTED					
TOTAL	50.785	44.526	31.887	44.365	50.874
DROUGHT	40.169	26.111	15.819	22.899	32.904
FLOOD	2.845	9.146	10.632	14.197	11.949
HAILSTORM	4.441	6.899	4.927	6.230	5.029
FROST	3.329	2.369	0.509	0.949	0.991
(2) AREAS HEAVILY AFFECTED (over 30% loss of crops)					
TOTAL	21.801	22.317	15.264	22.705	23.945
DROUGHT	17.969	12.485	7.015	10.063	15.303
FLOOD	0.924	5.025	5.395	8.949	6.128
HAILSTORM	2.026	3.630	2.654	3.353	1.989
FROST	0.879	1.177	0.201	0.341	0.525
(3) AREAS WHERE CROPS ARE COMPLETELY DESTROYED					
TOTAL	3.543	3.894	3.727	5.229	4.905
DROUGHT	3.337	N/A	1.219	2.049	2.745
FLOOD	0.139	N/A	1.979	2.689	1.799
HAILSTORM	0.056	N/A	0.508	0.485	0.279
FROST	0.011	N/A	0.021	0.007	0.083

N/A = Not available.

Unit: Million hectares.

Source: China Agriculture Yearbook, 1986 and 1989.

TABLE 2 OLS REGRESSION RESULTS

DEP VAR: GRAIN

	1952-1988		1952-1966		1970-1988	
VARIABLE	COEF.	STD. ERR.	COEF.	STD. ERR.	COEF.	STD. ERR.
INT	-1127.8	10501	-13568	5716	-15541	41987
LAND	186.92	70.66	240.96	38.69	291.51	290.5
MACHINERY	10.68	0.393	33.23	5.54	8.94	0.967
SHOUZAI	-114.63	25.52	-97.87	14.05	-152.95	65.63
CHENGZAI	-105.64	24.68	-54.29	17.52	-19.65	51.22
R-SQ	0.967		0.896		0.955	

DEP VAR: RICE

	1952-1988		1952-1966		1970-1988	
VARIABLE	COEF.	STD. ERR.	COEF.	STD. ERR.	COEF.	STD. ERR.
INT	-2205.3	6282.4	-9698.3	2886.7	6138	20194
LAND	98.26	42.27	130.35	19.54	39.35	139.7
MACHINERY	4.87	0.235	19.41	2.80	3.28	0.465
SHOUZAI	-61.91	15.27	-56.72	7.093	-47.73	31.57
CHENGZAI	-56.2	14.76	-22.61	8.84	-1.16	24.64
R-SQ	0.942		0.918		0.935	

DEP VAR: WHEAT

	1952-1988		1952-1966		1970-1988	
VARIABLE	COEF.	STD. ERR.	COEF.	STD. ERR.	COEF.	STD. ERR.
INT	-2282.4	2953	-4276.8	1754.6	-4137.8	18.81
LAND	35.17	19.89	46.07	11.87	50.84	125.1
MACHINERY	3.00	0.11	3.85	1.70	2.92	0.417
SHOUZAI	-16.49	7.18	-9.19	4.31	-36.3	28.26
CHENGZAI	-7.31	6.94	-5.34	5.38	4.52	22.06
R-SQ	0.968		0.650		0.938	

Note: Period 1952-1988 does not include 1967-1969.

TABLE 3 KALMAN FILTER MLE AND OLS REGRESSION RESULTS

TIME PERIOD: 1952-1988						
	MLE(1)		MLE(2)		OLS	
VARIABLE	COEF.	STD. ERR.	COEF.	STD. ERR.	COEF.	STD. ERR.
INT	-1458.7	14894	-1458.3	14992	-1127.8	10501
LAND	231.0	72.5	234.3	74.3	186.9	70.66
MACHINERY	9.578	0.378	9.37	0.377	10.68	0.393
SHOUZAI	-69.50	13.11	-68.67	12.96	-114.63	25.52
CHENGZAI	-30.30	6.02	-29.08	6.13	-105.64	24.68
1989 PRED.	41.29		41.31		35.20	
PRED. BIAS	0.53		0.55		5.56	

Note: Period 1952-1988 does not include 1967-1969.

TABLE 4 MLE AND OLS REGRESSION RESULTS

TIME PERIOD: 1970-1988				
	MLE		OLS	
VARIABLE	COEF.	STD. ERR.	COEF.	STD. ERR.
INT	-15541	41992	-15541	41987
LAND	291.46	189.8	291.51	290.5
MACHINERY	8.065	0.456	8.94	0.967
SHOUZAI	-152.97	65.26	-152.95	65.63
CHENGZAI	-19.67	49.86	-19.65	51.22
1989 PRED.	41.58		44.05	
PRED. BIAS	0.82		3.29	

The following was presented as a poster exhibit on the day of the conference:

Public School Enrollment Projections by State: Forecast Accuracy vs. Enrollment Size

Debra E. Gerald and DeeAnn Wright, National Center for Education Statistics

Abstract: The forecast accuracy of public school enrollment projections for the 50 States and the District of Columbia is demonstrated. Using data from 1970 to 1984, projections of public school enrollment are developed for 1985 through 1989. These projections are compared with actual public school enrollments during this period to calculate forecast errors. The mean absolute percentage error (MAPE) is used to measure forecast accuracy. The MAPEs by lead time are analyzed by the enrollment size of the state. Similarities and differences in patterns of errors among states are observed. It is expected that short-term projections are more accurate than long-term projections. Furthermore, states with the largest enrollment sizes are expected to have smaller MAPEs.

The following presentations were made on the day of the conference but papers were not submitted for publication in this document:

Using ARMAX Models in Practice

Jeffrey S. Butler, Internal Revenue Service

Abstract: Any single equation linear regression model can be derived from an ARMAX or Transfer Function model with appropriate restrictions on the lag polynomials. As a specific class of stochastic difference equations, ARMAX models allow for a much wider range of dynamics than the standard linear regression model, in terms of specifying both systematic and non-systematic components. There may be a tendency to avoid using ARMAX models in practice, however, as a result of difficulties in properly identifying the exact form of system dynamics prior to parameter estimation and forecasting. This paper discusses some of the advantages and disadvantages of ARMAX models, and provides selected simulation results which test the adequacy of ARMAX models against univariate ARIMA and multiple regression specifications using a minimum mean square error forecast criterion.

The Dutch Experience: Designing an Integrated Model to Assess the Greenhouse Effect

Michel den Elzen, Department of Global Biosphere, The Netherlands

Abstract: IMAGE is a mathematical simulation model designed to estimate the causes and consequences of the enhanced greenhouse effect. IMAGE was developed at the National Institute of Public Health and Environemntal Protection in The Netherlands. The model was commissioned by the Dutch government in order to help achieve four primary objectives: 1) To offer to government agencies a concise overview of the quantitative aspects of the greenhouse problem; 2) To increase awareness of the greenhouse problem among other sectors of society; 3) To provide insights into the likely effects of policy options concerning the enhanced greenhouse effect; and 4) To identify uncertainties or crucial gaps in current scientific knowledge. An overview of IMAGE and a brief computer demonstration will be given.

Mathematical Support of Integrated Environmental Models

O. J. Vrieze, University of Limburg, The Netherlands

Abstract: IMAGE is composed of a variety of submodels called modules. These modules are based on mathematical models describing physical-chemical processes and forecasts relevant to the greenhouse problem. At present these mathematical models and forecasts are relatively simple ones and need to be refined in order to better the understanding of the interaction of the various processes and to enhance the reliablility of IMAGE as a forecasting tool. The material will be presented in a manner understandable to nonspecialists.

The Australian Experience: Blood Bank Modeling

Roger D. Braddock, Griffith University, Australia

Abstract: A simulation model has been developed to model the behavior of blood in the hospital system. Blood and blood products are subject to supply and demand forces and also have a limited shelf life. The two main parameters are the ratio of blood returned to blood ordered, and the blood retention time in the sub-inventory. The model provides operational guidance on the level of a daily supply, in relation to these two important parameters and on the forecast demands for blood.

Analytical Tools to Increase Forecasting Models' Credibility

Jerzy A. Filar, University of Maryland, Baltimore County

Abstract: This talk will focus on the need to develop new analytical tools that will enhance the capabilities of environmental forecasting models. Emphasis will be placed on the need to improve the capabilities for variability/error analysis as well as the capabilities for "backward inference," and its policy analysis implications. It will be shown that even a sophisticated model such as IMAGE needs to improve these capabilities. Preliminary results of the variability analysis of IMAGE's forecasts will be reported.

☆U.S. GOVERNMENT PRINTING OFFICE: 1992-312-408/64639

NATIONAL AGRICULTURAL LIBRARY
1022352035

* NATIONAL AGRICULTURAL LIBRARY
1022352035